Sharp bounds on the MTE with sample selection

Possebom, Vitor

Yale University

30 October 2018

Sharp Bounds on the MTE with Sample Selection*

Vitor Possebom†
Yale University

First Draft: October 30th, 2018
Please click here for the most recent version

Abstract

I propose a Generalized Roy Model with sample selection that can be used to analyze treatment effects in a variety of empirical problems. First, I decompose, under a monotonicity assumption on the sample selection indicator, the MTR function for the observed outcome when treated as a weighted average of (i) the MTR on the outcome of interest for the always-observed sub-population and (ii) the MTE on the observed outcome for the observed-only-when-treated sub-population, and show that such decomposition can provide point-wise sharp bounds on the MTE of interest. I, then, show how to point-identify these bounds when the support of the propensity score is continuous. After that, I show how to (partially) identify the MTE of interest when the support of the propensity score is discrete.

Keywords: Marginal Treatment Effect, Sample Selection, Selection into Treatment, Partial Identification.

JEL Codes: C31, C35, C36

*I would like to thank Joseph Altonji, John Finlay, Yuichi Kitamura, Marianne Köhli, Helena Laneuville, Jaewon Lee, Yusuke Narita, Masayuki Sawada, Edward Vytlacil, Siuyuat Wong and seminar participants at Yale University for helpful suggestions.

†*Email:* vitoraugusto.possebom@yale.edu

1 Introduction

I propose a Generalized Roy Model (Heckman & Vytlacil 1999) with sample selection in which there is one outcome of interest that is observed only if the individual self-selects into the sample. So, in addition to the fundamental problem of causal analysis in which I only observe one of the potential outcomes due to endogenous self-selection into treatment, I also face a problem of endogenous sample selection. Such framework is useful to analyze many empirical problems: the effect of a job training program on wages (Heckman et al. (1999), Lee (2009), Chen & Flores (2015)), the college wage premium (Altonji (1993), Card (1999), Carneiro et al. (2011)), scarring effects (Heckman & Borjas (1980), Farber (1993), Jacobson et al. (1993)), the effect of an educational intervention on short- and long-term outcomes (Krueger & Whitmore (2001), Angrist et al. (2006), Angrist et al. (2009), Chetty et al. (2011), Dobbie & Jr. (2015)), the effect of a medical treatment on health quality (CASS (1984), U.S. Department of Health and Human Services (2004)), the effect of procedural laws on litigation outcomes (Helland & Yoon (2017)), and any randomized control trial that faces an attrition problem (DeMel et al. (2013), Angelucci et al. (2015)).

Under a monotonicity assumption on the sample selection indicator, I decompose the Marginal Treatment Response (MTR) function for the potential observed outcome when treated as a weighted average of (i) the MTR on the outcome of interest for the sub-population who is always observed and (ii) the Marginal Treatment Effect (MTE) on the observed outcome for the sub-population who is observed only when treated. Under a bounded (in one direction) support condition, such decomposition is useful because it allows me to propose point-wise sharp bounds on the MTE on the outcome of interest as a function of the MTR functions on the observed outcome, the maximum and (or) minimum of the support of the potential outcome, and the proportions of always-observed individuals and observed-only-when-treated individuals. I also show that it is impossible to construct bounds without extra assumptions when the support of the potential outcome is the entire real line.

I, then, proceed to show that those bounds are well-identified. When the support of the propensity score is an interval, the relevant objects are point-identified by applying the local instrumental variable approach (LIV, see Heckman & Vytlacil (1999)) to the expectations of the observed outcome and of the selection indicator conditional on the propensity score and the treatment status. However, in many empirical applications, the support of the propensity score is a finite set. In such context, I can identify bounds on the MTE of interest by adapting the nonparametric bounds proposed by Mogstad et al. (2017) or the flexible parametric approach suggested by Brinch et al. (2017) to encompass a sample selection problem. When using the nonparametric approach, the bounds on the MTE of interest are simply an outer set that contains the true MTE, i.e., they are not point-wise sharp anymore.

Partial identification of the MTE of interest is useful for two reasons. First, bounds on the MTE can be used to shed light on the heterogeneity of treatment effects, allowing the researcher to understand who benefits and who loses with a specific treatment. Such knowledge can be used to optimally design policies that provide incentives to agents to take a treatment. Second, bounds on the MTE can be used to construct bounds in any treatment effect parameter that is written as a weighted integral of the MTE. For example, by taking a weighted average of the point-wise sharp bounds on the MTE, one can bound the average treatment effect (ATE), the average treatment effect on the treated (ATT), any local average treatment effect (LATE, Imbens & Angrist (1994)) and any policy-relevant treatment effect (PRTE, Heckman & Vytlacil (2001*b*)). Although such bounds may not be sharp for any specific parameter, they are a general and easy-to-apply solution to many empirical problems. Therefore, if the applied researcher is interested in a parameter that already has specific bounds for it (e.g., intention-to-treat treatment effect (ITT) by Lee (2009) and LATE by Chen & Flores (2015)), he or she should use a specialized tool. However, if the applied researcher is interested in parameters without specialized bounds (e.g., the PRTE or LATEs outside the support of the propensity score), he or she may take a weighted integral of point-wise sharp bounds on the MTE of interest. In other words, facing a trade-off between empirical flexibility and sharpness, the partial identification tool proposed in this paper focus on empirical flexibility while still ensuring some notion of sharpness.

I make contributions to two literatures: identification of treatment effects using an instrument and identification of treatment effects with sample selection.

The literature about treatment effects with an instrument is enormous and I only briefly discuss it. Imbens & Angrist (1994) show that we can identify the Average Treatment Effect for the Compliers (LATE). Heckman & Vytlacil (1999), Heckman & Vytlacil (2005) and Heckman et al. (2006) define the MTE and explain how to compute any treatment effect as a weighted average of the MTE. However, if the support of the propensity score is not the unit interval, then it is not possible to recover some important treatment effects, such as the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATT) and the Policy Relevant Treatment Effect (PRTE, Heckman & Vytlacil (2001*b*)). A parametric solution to this problem is given by Brinch et al. (2017), who identify a flexible polynomial function for the MTE whose degree is defined by the cardinality of the propensity score support.

A nonparametric solution to the impossibility of identifying the ATE and the ATT is bounding them. Mogstad et al. (2017) use the information contained on IV-like estimands to construct non-parametrically worst- and best- case bounds on policy-relevant treatment effects. Other authors focus on imposing weak monotonicity assumptions or a structural model. In the first group, Manski (1990), Manski (1997) and Manski & Pepper (2000) propose bounds for the ATE and ATT. Chen et al. (2017) propose an average monotonicity condition combined with a mean dominance condition across subpopulation groups and sharpen the bounds previously proposed. Huber et al. (2017) add a mean independence condition within subpopulation groups and bound not only the ATE and ATT when there is noncompliance, but also the Average Treatment Effect on the Untreated (ATUT) and the ATE for always-takers and never-takers (ATE-AT and ATE-NT).

Complementing the weak monotonicity approach, the structural approach has focused mainly on the binary outcome case due to the need to impose bounded outcome variables. Heckman & Vytlacil (2001*a*), Bhattacharya et al. (2008), Chesher (2010), Chiburis (2010), Shaikh & Vytlacil (2011) and Bhattacharya et al. (2012) made important contributions to this literature, bounding the ATE and the ATT. While Bhattacharya et al. (2008), Shaikh &

Vytlacil (2011) and Bhattacharya et al. (2012) consider a thresholding crossing model on the treatment and the outcome variable, Chiburis (2010) assumes a thresholding crossing model only on the outcome variable.

I contribute to this literature about identifying treatment effects using an instrument by extending the non-parametric approach by Mogstad et al. (2017) and the flexible parametric approach by Brinch et al. (2017) to encompass a sample-selection problem. By doing so, I can partially identify the MTE function on the outcome of interest instead of on the observed outcome.

The literature about identification of treatment effects with sample selection is vast and I only briefly discuss it. The control function approach is a possible solution to it and is analyzed by Heckman (1979), Ahn & Powell (1993) and Newey et al. (1999), encompassing parametric, semiparametric and nonparametric tools. Using auxiliary data is another possible solution and is studied by Chen et al. (2008). A nonparametric solution that requires weaker conditions is bounding. In a seminal paper, Lee (2009) imposes a weak monotonicity assumption on the relationship between sample selection and treatment assignment to sharply bound the Intention-to-Treat Average Treatment Effect (ITT) for the subpopulation of always-observed individuals. Using techniques developed by Frangakis & Rubin (2002), Blundell et al. (2007) and Imai (2008) and a weak monotonicity assumption, Blanco et al. (2013) bound the Intention-to-Treat Quantile Treatment Effect (Q-ITT) for the always-observed individuals. Moreover, by imposing weak dominance assumptions across subpopulation groups, they can sharpen the ITT bounds proposed by Lee (2009). Huber & Mellace (2015) additionally impose a bounded support for the outcome variable and propose bounds on the ITT for two other subpopulations: observed-only-when-treated individuals, and observed-only-when-untreated individuals. Complementary to those studies, Lechner & Mell (2010) derive bounds for the ITT and the Q-ITT for the treated-and-observed subpopulation, Mealli & Pacini (2013) derive bounds for the ITT when the exclusion restriction is violated and there are two outcome variables, and Behaghel et al. (2015) combines techniques developed by Heckman (1979) and Lee (2009) to propose bounds for the ATE in a survey framework in which the interviewer tries to contact the surveyed individual multiple times.

In the intersection of both literatures, a few authors address the problem of sample selection and endogenous treatment simultaneously. Huber (2014) point-identifies the ATE and the Quantile Treatment Effect (QTE) for the observed sub-population and for the entire population using a nested propensity score based on a instrument for sample selection. Fricke et al. (2015), by using a random treatment assignment and a continuous exogenous variable to instrument for treatment status and sample selection, point-identify the LATE. Lee & Salanie (2016), who also include sample selection in a Generalized Roy Model, use two continuous instruments to provide control functions for the selection into treatment and sample selection problems, allowing them to point-identify the MTE.

Although the three previous contributions are important, finding a credible instrument for sample selection is hard, especially in Labor Economics. For this reason, it is important to develop tools that do not rely on the existence of an instrument for sample selection. Frolich & Huber (2014) point-identify the unconditional LATE under an conditional IV independence assumption and a predetermined sample-selection assumption, ruling out an contemporaneous relationship between the potential outcomes and the sample selection problem. Chen & Flores (2015) derive bounds for Average Treatment Effect for the always-observed compliers (LATE-OO) by combining one instrument with a double exclusion restriction with monotonicity assumptions on the sample selection and the selection into treatment problems. Moreover, Blanco et al. (2017) and Steinmayr (2014) extend the work by Chen & Flores (2015) by, respectively, considering a censored outcome variable and analyzing mixture variables combining four strata.

I contribute to the literature about identification of treatment effects with sample selection by partially identifying the MTE on the always-observed subsample allowing for an contemporaneous relationship between the potential outcomes and the sample selection problem, and using only one (discrete) instrument combined with a monotonicity assumption. Doing so is theoretically important, because it can unify, in one framework, the bounds for different treatment effects with sample selection, and empirically relevant, because it allow us to partially identify any treatment effect on the outcome of interest in many empirical problems.

This paper proceeds as follows: section 2 details the Generalized Roy Model with sample selection; section 3 explains how to derive bounds for the MTE of interest; and sections 4 and 5 discuss identification of the MTE bounds when the support of the propensity score is continuous or discrete. Finally, section 6 discuss further work.

2 Framework

I begin with the classical potential outcome framework by Rubin (1974) and modify it to include a sample selection problem. Let Z be an instrumental variable whose support is given by $\mathcal{Z}$, X be a vector of covariates whose support is given by $\mathcal{X}$, $W := (X, Z)$ be a vector that combines the covariates and the instrument whose support is given by $\mathcal{W} := \mathcal{X} \times \mathcal{Z}$, D be a treatment status indicator, Y_0^* be the potential outcome of interest when the person is not treated, and Y_1^* be the potential outcome of interest when the person is treated. The outcome variable of interest (e.g., wages) is $Y^* := D \cdot Y_1^* + (1 - D) \cdot Y_0^*$. Moreover, let S_1 and S_0 be potential sample selection indicators when treated and when not treated, and define $S := D \cdot S_1 + (1 - D) \cdot S_0$ as the sample selection indicator. Define $Y := S \cdot Y^*$ as the observed outcome (e.g., labor earnings). I also define $Y_1 := S_1 \cdot Y_1^*$ and $Y_0 := S_0 \cdot Y_0^*$ as the potential observable outcomes. Observe that, following Lee (2009) and Chen & Flores (2015), my notation implicit imposes two exclusion restrictions: Z has no direct impact on the potential outcome of interest nor on the sample selection indicator. The second exclusion restriction requires attention in empirical applications. On the one hand, it may be a strong assumption in randomized control trials if sample selection is due to attrition and initial assignment has an effect on the subject's willingness to contact the researchers. On the other hand, it may be a reasonable assumption in many labor market applications in which initial random assignment to a job training program has no impact on future employment status.

I model sample selection and selection into treatment using the Generalized Roy Model (Heckman & Vytlacil 1999). Let U and V be random variables, and $P : \mathcal{W} \to \mathbb{R}$ and

$Q : \{0,1\} \times \mathcal{X} \to \mathbb{R}$ be unknown functions. I assume that:

$$D \coloneqq \mathbf{1}\{P(W) \geq U\} \tag{1}$$

and

$$S \coloneqq \mathbf{1}\{Q(D,X) \geq V\}. \tag{2}$$

As Vytlacil (2002) shows, equations (1) and (2) are equivalent to assuming monotonicity conditions on the selection into treatment problem (Imbens & Angrist (1994)) and on the sample selection problem (Lee (2009)). I stress that both monotonicity assumptions are testable using the tools developed by Machado et al. (2018). Note also that, given equation (2), $S_0 = \mathbf{1}\{Q(0,X) \geq V\}$ and $S_1 = \mathbf{1}\{Q(1,X) \geq V\}$.

The random variables U and V are jointly continuously distributed conditional on X with density $f_{U,V|X} : \mathbb{R}^2 \times \mathcal{X} \to \mathbb{R}$ and cumulative distribution function $F_{U,V|X} : \mathbb{R}^2 \times \mathcal{X} \to \mathbb{R}$. As is well known in the literature, equations (1) and (2) can be rewritten as

$$\begin{aligned} D &= \mathbf{1}\left\{F_{U|X}(P(W)|X) \geq F_{U|X}(U|X)\right\} = \mathbf{1}\left\{\tilde{P}(W) \geq \tilde{U}\right\} \\ S &= \mathbf{1}\left\{F_{V|X}(Q(D,X)|X) \geq F_{V|X}(V|X)\right\} = \mathbf{1}\left\{\tilde{Q}(D,X) \geq \tilde{V}\right\} \end{aligned}$$

where $\tilde{P}(W) \coloneqq F_{U|X}(P(W)|X)$, $\tilde{U} \coloneqq F_{U|X}(U|X)$, $\tilde{Q}(D,X) \coloneqq F_{V|X}(Q(D,X)|X)$, and $\tilde{V} \coloneqq F_{V|X}(V|X)$. Consequently, the marginal distributions of $\tilde{U}$ and $\tilde{V}$ conditional on X follow the standard uniform distribution. Since this is merely a normalization, I drop the tilde and mantain throughout the paper the normalization that the marginal distributions of U and V conditional on X follow the standard uniform distribution and that $(P(w), Q(d,x)) \in [0,1]^2$ for any $(x,z,d) \in \mathcal{W} \times \{0,1\}$. I also assume that:

Assumption 1 *The instrument Z is independent of all latent variables given the covariates X, i.e., $Z \perp\!\!\!\perp (U, V, Y_0^*, Y_1^*) | X$.*

Assumption 2 *The distribution of $P(W)$ given X is nondegenerate.*

Assumption 3 *The first and second population moments of the counterfactual variables are finite, i.e.,* $\mathbb{E}\left[|Y_d^*|\right] < +\infty$, $\mathbb{E}\left[(Y_d^*)^2\right] < +\infty$, *and* $\mathbb{E}\left[|S_d|\right] < +\infty$ *for any* $d \in \{0, 1\}$.

Assumption 4 *Both treatment groups exist for any value of* X, *i.e.,* $0 < \mathbb{P}\left[D = 1 \,|X\,\right] < 1$.

Assumption 5 *The covariates* X *are invariant to counterfactual manipulations, i.e.,* $X_0 = X_1 = X$, *where* X_0 *and* X_1 *are the counterfactual values of* X *that would be observed when the person is, respectively, not treated or treated.*

Assumption 6 *The potential outcomes* Y_0^* *and* Y_1^* *have the same support, i.e.,* $\mathcal{Y}^* := \mathcal{Y}_0^* = \mathcal{Y}_1^*$, *where* $\mathcal{Y}_0^* \subseteq \mathbb{R}$ *is the support of* Y_0^* *and* $\mathcal{Y}_1^* \subseteq \mathbb{R}$ *is the support of* Y_1^*.

Assumption 7 *Define* $\underline{y}^* := \inf\{y \in \mathcal{Y}^*\} \in \mathbb{R} \cup \{-\infty\}$ *and* $\overline{y}^* := \sup\{y \in \mathcal{Y}^*\} \in \mathbb{R} \cup \{\infty\}$. *I assume that* $\underline{y}^*$ *and* $\overline{y}^*$ *are known, and that*

1. $\underline{y}^* > -\infty$, $\overline{y}^* = \infty$ *and* $\mathcal{Y}^*$ *is an interval, or*
2. $\underline{y}^* = -\infty$, $\overline{y}^* < \infty$ *and* $\mathcal{Y}^*$ *is an interval, or*
3. $\underline{y}^* > -\infty$, $\overline{y}^* < \infty$ *and*
   (a) $\mathcal{Y}^*$ *is an interval or*
   (b) $\underline{y}^* \in \mathcal{Y}^*$ *and* $\overline{y}^* \in \mathcal{Y}^*$.

I stress that assumption 7 is fairly general. Case 1 covers continuous random variables whose support is convex and bounded below (e.g.: wages), while Case 3.a covers continuous variables with bounded convex support (e.g.: test scores). Case 3.b encompasses not only binary variables, but also any discrete variable whose support is finite (e.g.: years of education). It also includes mixed random variables whose support is not an interval but achieves its maximum and minimum. I also highlight that proposition 12 shows that assumption 7 is partially necessary to the existence of bounds on the MTE of interest in the sense that, if $\underline{y}^* = -\infty$ and $\overline{y}^* = +\infty$, then it is impossible to bound the marginal treatment effect on the outcome of interest without any extra assumption.

Assumption 8 *Treatment has a positive effect on the sample selection indicator for all individuals, i.e.,* $Q(1,x) > Q(0,x) > 0$ *for any* $x \in \mathcal{X}$.

Assumption 8 goes beyond the monotonicity condition implicitly imposed by equation (2) by assuming that the direction of the effect of treatment on the sample selection indicator is known and positive, i.e., $Q(1,x) \geq Q(0,x)$ for any $x \in \mathcal{X}$. In this sense, it is a standard assumption in the literature.[1] Most importantly, it is also a testable assumption using the tools developed by Machado et al. (2018), because, under monotone sample selection (equation (2)), identification of the sign of the ATE on the selection indicator provides a test for assumption 8. However, assumption 8 is slightly stronger than what is usually imposed in the literature, because it additionally imposes $Q(0,x) > 0$ and $Q(1,x) > Q(0,x)$ for any $x \in \mathcal{X}$. I stress that the first inequality implies that there is a sub-population who is always observed, allowing me to properly define my target parameter — the marginal treatment effect on the outcome of interest for the always-observed population. I also highlight that the second inequality implies that there is a sub-population who is observed only when treated, making the problem theoretically interesting by eliminating trivial cases of point-identification of the MTE of interest as discussed in proposition 9. Finally, I emphasize that all my results can be stated and derived with some obvious changes if I impose $Q(0,x) > Q(1,x) > 0$ for any $x \in \mathcal{X}$ instead of assumption 8, as it is done in Appendix E. I also discuss an agnostic approach to monotonicity in Appendix F.

3 Bounds on the MTE on the outcome of interest

The target parameter, the MTE on the outcome of interest for the sub-population who is always observed, is given by

$$\begin{aligned}\Delta_{Y^*}^{OO}(x,u) &\coloneqq \mathbb{E}\left[Y_1^* - Y_0^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \\ &= \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] - \mathbb{E}\left[Y_0^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right]\end{aligned} \tag{3}$$

[1] Lee (2009) and Chen & Flores (2015) write it in an equivalent way as $S_1 \geq S_0$.

for any $u \in [0,1]$ and any $x \in \mathcal{X}$, and is a natural parameter of interest. In labor market applications where sample selection is due to observing wages only when agents are employed, it is the effect on wages for the subpopulation who is always employed. In medical applications where sample selection is due to the death of a patient, it is the effect on health quality for the subpopulation who survives regardless of the treatment status. In the education literature where sample selection is due to students quiting school, it is the effect on test scores for the subpopulation who do not drop out of school regardless of the treatment status. In all those cases, the target parameter captures the intensive margin of the treatment effect.[2]

Other possibly interesting parameters are the MTE on the outcome of interest for the sub-population who is never observed ($\mathbb{E}\left[Y_1^* - Y_0^* \left| X = x, U = u, S_0 = 0, S_1 = 0\right.\right]$), the MTR function under no treatment for the outcome of interest for the sub-population who is observed only when treated ($\mathbb{E}\left[Y_0^* \left| X = x, U = u, S_0 = 0, S_1 = 1\right.\right]$) and MTR function under treatment for the outcome of interest for the sub-population who is observed only when treated ($\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 0, S_1 = 1\right.\right]$). While the last parameter can be partially identified (Appendix D), the first two parameters are impossible to point-identify or bound in a informative way because the outcome of interest (Y_0^* or Y_1^*) is never observed for the conditioning sub-populations. Note also that the sub-population who is observed only when not treated ($S_0 = 1$ and $S_1 = 0$) do not exist by assumption 8.

We, now, focus on the target parameter $\Delta_{Y^*}^{OO}(x,u)$ given by equation (3). The second right-hand term in equation (3) can be written as[3]

$$\mathbb{E}\left[Y_0^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] = \frac{m_0^Y(x,u)}{m_0^S(x,u)}, \tag{4}$$

where I define $m_0^Y(x,u) := \mathbb{E}\left[Y_0 \left| X = x, U = u\right.\right]$ and $m_0^S(x,u) := \mathbb{E}\left[S_0 \left| X = x, U = u\right.\right]$ as the MTR functions associated to the counterfactual variables Y_0 and S_0 respectively. In this section, I assume that all terms in the right-hand side of equation (4) are point-identified,

[2] If the researcher is interested in the extensive margin of the treatment effect, captured by the MTE on the observable outcome ($\mathbb{E}\left[Y_1 - Y_0 \left| X = x, U = u\right.\right]$) and by the MTE on the selection indicator ($\mathbb{E}\left[S_1 - S_0 \left| X = x, U = u\right.\right]$), he or she can apply the identification strategies described by Heckman et al. (2006), Brinch et al. (2017) and Mogstad et al. (2017).

[3] Appendix A.1 contains a proof of this claim.

postponing the discussion about their identification to sections 4 and 5.

The first right-hand term in equation (3) can be written as[4]

$$\mathbb{E}\left[Y_1^*\,|X=x,U=u,S_0=1,S_1=1\right]=\frac{m_1^Y\left(x,u\right)-\Delta_Y^{NO}\left(x,u\right)\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}, \tag{5}$$

where $m_1^Y\left(x,u\right)\coloneqq\mathbb{E}\left[Y_1\,|X=x,U=p\right]$ is the MTR function associated to the counterfactual variable Y_1, $\Delta_Y^{NO}\left(x,u\right)\coloneqq\mathbb{E}\left[Y_1-Y_0\,|X=x,U=u,S_0=0,S_1=1\right]$ is the MTE on the observed outcome Y for the sub-population who is observed only when treated, $\Delta_S\left(x,u\right)\coloneqq\mathbb{E}\left[S_1-S_0\,|X=x,U=u\right]=m_1^S\left(x,u\right)-m_0^S\left(x,u\right)$ is the MTE on the selection indicator, and $m_1^S\left(x,u\right)\coloneqq\mathbb{E}\left[S_1\,|X=x,U=u\right]$ is the MTR function associated to the counterfactual variable S_1. In this section, I also assume that $m_1^Y\left(x,u\right)$ and $\Delta_S\left(x,u\right)$ are point-identified, postponing the discussion about their identification to sections 4 and 5.

Although point-identification of $\mathbb{E}\left[Y_1^*\,|X=x,U=u,S_0=1,S_1=1\right]$ is not possible, I can find identifiable bounds for it.[5]

Proposition 9 *Suppose that $m_0^Y\left(x,u\right)$, $m_1^Y\left(x,u\right)$, $m_0^S\left(x,u\right)$ and $\Delta_S\left(x,u\right)$ are point-identified.*

Under assumptions 1-6, 7.1 and 8, the bounds on $\mathbb{E}\left[Y_1^*\,|X=x,U=u,S_0=1,S_1=1\right]$ are given by

$$\underline{y}^*\leq\mathbb{E}\left[Y_1^*\,|X=x,U=u,S_0=1,S_1=1\right]\leq\frac{m_1^Y\left(x,u\right)-\underline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}. \tag{6}$$

Under assumptions 1-6, 7.2 and 8, the bounds on $\mathbb{E}\left[Y_1^*\,|X=x,U=u,S_0=1,S_1=1\right]$ are given by

$$\frac{m_1^Y\left(x,u\right)-\overline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}\leq\mathbb{E}\left[Y_1^*\,|X=x,U=u,S_0=1,S_1=1\right]\leq\overline{y}^*. \tag{7}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)) and 8, the bounds on

[4] Appendix A.2 contains a proof of this claim.

[5] Appendix A.3 contains a proof of this proposition.

$\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right]$ *are given by*

$$\frac{m_1^Y(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} \leq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)}. \tag{8}$$

There is a important remark to be made about the bounds of proposition 9. Note that, even when the support is bounded in only one direction (assumptions 7.1 and 7.2), it is possible to derive lower and upper bounds on $\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right]$.

At this point, it is also important to understand the determinants of the width of those bounds. First, if there is no sample selection problem at all ($\mathbb{P}\left[S_0 = 1, S_1 = 1 \left| X = x, U = u\right.\right] = 1$), then $m_0^S(x,u) = 1$, $\Delta_S(x,u) = 0$, implying tighter bounds in equations (6) and (7) and point-identification in equation (8). Second and most importantly, if there is no problem of differential sample selection with respect to treatment status ($\mathbb{P}\left[S_0 = 0, S_1 = 1 \left| X = x, U = u\right.\right] = 0$), then $\Delta_S(x,u) = 0$, once more implying tighter bounds in equations (6) and (7) and point-identification in equation (8). Both cases are theoretically uninteresting and ruled out by assumption 8.

Finally, combining equations (3) and (4) and proposition 9, I can partially identify the target parameter $\Delta_{Y^*}^{OO}(x,u)$:

Corollary 10 *Suppose that $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$ and $\Delta_S(x,u)$ are point-identified.*

Under assumptions 1-6, 7.1 and 8, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ are given by

$$\Delta_{Y^*}^{OO}(x,u) \geq \underline{y}^* - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{9}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{10}$$

Under assumptions 1-6, 7.2 and 8, the bounds on $\Delta^{OO}_{Y^*}(x,u)$ are given by

$$\Delta^{OO}_{Y^*}(x,u) \geq \frac{m^Y_1(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m^S_0(x,u)} - \frac{m^Y_0(x,u)}{m^S_0(x,u)} =: \underline{\Delta^{OO}_{Y^*}}(x,u) \tag{11}$$

and that

$$\Delta^{OO}_{Y^*}(x,u) \leq \overline{y}^* - \frac{m^Y_0(x,u)}{m^S_0(x,u)} =: \overline{\Delta^{OO}_{Y^*}}(x,u). \tag{12}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)) and 8, the bounds on $\Delta^{OO}_{Y^*}(x,u)$ are given by

$$\Delta^{OO}_{Y^*}(x,u) \geq \max\left\{\frac{m^Y_1(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m^S_0(x,u)}, \underline{y}^*\right\} - \frac{m^Y_0(x,u)}{m^S_0(x,u)} =: \underline{\Delta^{OO}_{Y^*}}(x,u) \tag{13}$$

and that

$$\Delta^{OO}_{Y^*}(x,u) \leq \min\left\{\frac{m^Y_1(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m^S_0(x,u)}, \overline{y}^*\right\} - \frac{m^Y_0(x,u)}{m^S_0(x,u)} =: \overline{\Delta^{OO}_{Y^*}}(x,u). \tag{14}$$

Most importantly, I can show that[6]:

Proposition 11 *Suppose that the functions m^Y_0, m^Y_1, m^S_0 and Δ_S are point-identified at every pair $(x,u) \in \mathcal{X} \times [0,1]$. Under assumptions 1-6, 7 (sub-cases 1, 2, 3(a) or 3(b)) and 8, the bounds $\underline{\Delta^{OO}_{Y^*}}$ and $\overline{\Delta^{OO}_{Y^*}}$, given by corollary 10, are point-wise sharp, i.e., for any $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \left(\underline{\Delta^{OO}_{Y^*}}(\overline{x},\overline{u}), \overline{\Delta^{OO}_{Y^*}}(\overline{x},\overline{u})\right)$, there exist random variables $\left(\tilde{Y}^*_0, \tilde{Y}^*_1, \tilde{U}, \tilde{V}\right)$ such that*

$$\Delta^{OO}_{\tilde{Y}^*}(\overline{x},\overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}^*_1 - \tilde{Y}^*_0 \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x},\overline{u}), \tag{15}$$

$$\mathbb{P}\left[\left(\tilde{Y}^*_0, \tilde{Y}^*_1, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{16}$$

and

$$F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) = F_{Y,D,S,Z,X}(y,d,s,z,\overline{x}) \tag{17}$$

[6] Appendix A.4 contains the proof of this proposition. Note that, if the functions m^Y_0, m^Y_1, m^S_0 and Δ_S are point-identified only in a subset of the unit interval, then point-wise sharpness holds only in that subset.

for any $(y, d, s, z) \in \mathbb{R}^4$*, where* $\tilde{D} := \mathbf{1}\left\{P(X, Z) \geq \tilde{U}\right\}$*,* $\tilde{S}_0 = \mathbf{1}\left\{Q(0, X) \geq \tilde{V}\right\}$*,* $\tilde{S}_1 = \mathbf{1}\left\{Q(1, X) \geq \tilde{V}\right\}$*,* $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$*,* $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$*.*

Intuitively, proposition 11 says that, for any $\delta(\overline{x}, \overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x}, \overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x}, \overline{u})\right)$, it is possible to create candidate random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ that generate the candidate marginal treatment effect $\delta(\overline{x}, \overline{u})$, satisfy the bounded support condition — a restriction imposed by my model (assumption 7) and summarized in equation (16) — and generate the same distribution of the observable variables — a restriction imposed by the data and summarized in equation (17). In other words, the data and the model in section 2 do not generate enough restrictions to refute that the true target parameter $\Delta_{\tilde{Y}^*}^{OO}(\overline{x}, \overline{u})$ is equal to the candidate target parameter $\delta(\overline{x}, \overline{u})$.

Moreover, the bounded support condition (assumption 7) is partially necessary to the existence of bounds on the target parameter $\Delta_{\tilde{Y}^*}^{OO}(\overline{x}, \overline{u})$. When the support is unbounded in both directions (i.e., $\underline{y}^* = -\infty$ and $\overline{y}^* = +\infty$), then it is impossible to derive bounds on the target parameter $\Delta_{\tilde{Y}^*}^{OO}(\overline{x}, \overline{u})$ without any extra assumption. Proposition 12 formalizes this last statement.[7]

Proposition 12 *Suppose that the functions* m_0^Y*,* m_1^Y*,* m_0^S *and* Δ_S *are point-identified at every pair* $(x, u) \in \mathcal{X} \times [0, 1]$*. Impose assumptions 1-6 and 8. If* $\mathcal{Y}^* = \mathbb{R}$*, then, for any* $\overline{u} \in [0, 1]$*,* $\overline{x} \in \mathcal{X}$ *and* $\delta(\overline{x}, \overline{u}) \in \mathbb{R}$*, there exist random variables* $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ *such that*

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x}, \overline{u}) := \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x}, \overline{u}), \tag{18}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0, 1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0, 1], \tag{19}$$

and

$$F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z, X}(y, d, s, z, \overline{x}) = F_{Y, D, S, Z, X}(y, d, s, z, \overline{x}) \tag{20}$$

for any $(y, d, s, z) \in \mathbb{R}^4$*, where* $\tilde{D} := \mathbf{1}\left\{P(X, Z) \geq \tilde{U}\right\}$*,* $\tilde{S}_0 = \mathbf{1}\left\{Q(0, X) \geq \tilde{V}\right\}$*,* $\tilde{S}_1 = \mathbf{1}\left\{Q(1, X) \geq \tilde{V}\right\}$*,* $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$*,* $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$*.*

[7] Appendix A.5 contains the proof of this proposition.

This impossibility result is interesting in light of the previous literature about partial identification of treatment effects with sample selection. In the case of the ITT (Lee (2009)) and the LATE (Chen & Flores (2015)), it is possible to construct sharp and informative bounds even when the support of the potential outcome is the entire real line. However, when focusing on a specific point of the MTE function, it is impossible to construct informative bounds when $\mathcal{Y}^* = \mathbb{R}$ due to the local nature of the target parameter.

There are two important remarks about the results I have just derived. First, the bounds under assumption 7 (corollary 10) are not only sharp, but can also be informative as the numerical example in appendix B and the empirical application in section **??** illustrate. Second, propositions 11 and 12 do not impose any smoothness condition on the joint distribution of $(Y_0^*, Y_1^*, U, V, Z, X)$. In particular, the conditional cumulative distribution functions $F_{V|X,U}$, $F_{Y_0^*|X,U,V}$ and $F_{Y_1^*|X,U,V}$ are allowed to be discontinuous functions of U at the point $\overline{u}$. Appendix G states and proves a sharpness result similar to proposition 11 and an impossibility result similar to proposition 12 when $F_{V|X,U}$, $F_{Y_0^*|X,U,V}$ and $F_{Y_1^*|X,U,V}$ must be continuous functions of U.

Now, it is important to discuss the empirical relevance of partially identifying the MTE of interest. First, bounds on the MTE can illuminate the heterogeneity of the treatment effect, allowing the researcher to understand who benefits and who loses with a specific treatment. This is important because common parameters (e.g.: ATE, ATT, LATE) can be positive even when most people lose with a policy if the few winners have very large gains. Moreover, knowing, even partially, the MTE function can be useful to optimally design policies that provides incentives to agents to take some treatment. Second, I can use the MTE bounds to partially identify any treatment effect that is described as a weighted integral of $\Delta_{Y^*}^{OO}(x,u)$ because

$$\begin{aligned}\int_0^1 \left(\underline{\Delta_{Y^*}^{OO}}(x,u)\right) \cdot \omega(x,u)\, \mathrm{d}u &\leq \int_0^1 \Delta_{Y^*}^{OO}(x,u) \cdot \omega(x,u)\, \mathrm{d}u \\ &\leq \int_0^1 \left(\overline{\Delta_{Y^*}^{OO}}(x,u)\right) \cdot \omega(x,u)\, \mathrm{d}u, \end{aligned} \tag{21}$$

where $\omega(x,\cdot)$ is a weighting function. Even though such bounds may not be sharp for any specific parameter, they are a general and off-the-shelf solution to many empirical problems. As a consequence of this trade-off, I recommend the applied researcher to use a specialized tool if he or she is interested in a parameter that already has specific bounds for it (e.g., intention-to-treat treatment effect (ITT) by Lee (2009) and LATE by Chen & Flores (2015)). However, I suggest the applied research to easily compute a weighted integral of point-wise sharp bounds on the MTE of interest if he or she is interested in parameters without specialized bounds (e.g., the PRTE or LATEs outside the support of the propensity score). In other words, facing a trade-off between empirical flexibility and sharpness, the partial identification tool proposed in this paper focus on empirical flexibility while still ensuring point-wise sharpness of the bounds on the MTE of interest.

Tables 1 and 2 show some of the treatment effect parameters that can be partially identified using inequality (21). More examples are given by Heckman et al. (2006, Tables 1A and 1B) and Mogstad et al. (2017, Table 1).

Table 1: Treatment Effects as Weighted Integrals of the Marginal Treatment Effect

$ATE^{OO} = \mathbb{E}\left[Y_1^* - Y_0^* \left\vert S_0 = 1, S_1 = 1\right.\right] = \int_0^1 \Delta_{Y^*}^{OO}(u)\, \mathrm{d}u$
$ATT^{OO} = \mathbb{E}\left[Y_1^* - Y_0^* \left\vert D = 1, S_0 = 1, S_1 = 1\right.\right] = \int_0^1 \Delta_{Y^*}^{OO}(u) \cdot \omega_{ATT}(u)\, \mathrm{d}u$
$ATU^{OO} = \mathbb{E}\left[Y_1^* - Y_0^* \left\vert D = 0, S_0 = 1, S_1 = 1\right.\right] = \int_0^1 \Delta_{Y^*}^{OO}(u) \cdot \omega_{ATU}(u)\, \mathrm{d}u$
$LATE^{OO}(\underline{u},\overline{u}) = \mathbb{E}\left[Y_1^* - Y_0^* \left\vert U \in [\underline{u},\overline{u}], S_0 = 1, S_1 = 1\right.\right] = \int_0^1 \Delta_{Y^*}^{OO}(u) \cdot \omega_{LATE}(u)\, \mathrm{d}u$

Source: Heckman et al. (2006) and Mogstad et al. (2017). Note: Conditioning on X is kept implicit in this table for brevity.

In appendix C, I impose an extra mean-dominance assumption and sharpen the bounds given by corollary 10.

Table 2: Weights

$$\omega_{ATT}(x,u) = \frac{\int_u^1 f_{P(W)|X}(p\,|x)\,\mathrm{d}p}{\mathbb{E}[P(W)\,|X=x]}$$

$$\omega_{ATU}(x,u) == \frac{\int_0^u f_{P(W)|X}(p\,|x)\,\mathrm{d}p}{1-\mathbb{E}[P(W)\,|X=x]}$$

$$\omega_{LATE}(x,u) = \frac{\mathbf{1}\{u \in [\underline{u}, \overline{u}]\}}{\overline{u}-\underline{u}}$$

Source: Heckman et al. (2006) and Mogstad et al. (2017). Note: Conditioning on X is kept implicit in this table for brevity and $f_{P(W)|X}$ is the density of the propensity score given X.

4 Partial identification when the support of the propensity score is an interval

Here, I fix $x \in \mathcal{X}$ and impose that the support of the propensity score, defined by $\mathcal{P}_x \coloneqq \{P(x,z) : z \in \mathcal{Z}\}$, is an interval[8]. Then, under assumptions 1-5, the MTR functions associated to any variable $A \in \{Y, S\}$ are point-identified by[9]:

$$m_0^A(x,p) = \mathbb{E}[A\,|X=x, P(W)=p, D=0] - \frac{\partial \mathbb{E}[A\,|X=x, P(W)=p, D=0]}{\partial p} \cdot (1-p), \tag{22}$$

and

$$m_1^A(x,p) = \mathbb{E}[A\,|X=x, P(W)=p, D=1] + \frac{\partial \mathbb{E}[A\,|X=x, P(W)=p, D=1]}{\partial p} \cdot p \tag{23}$$

for any $p \in \mathcal{P}_x$.

Finally, the point-wise sharp bounds on $\Delta_{Y^*}^{OO}(x,p)$ are point-identified by combining corollary 10, equations (22) and (23), and the fact that $\Delta_S(x,p) = m_1^S(x,p) - m_0^S(x,p)$.

[8] $\mathcal{P}_x$ as an interval may be achieved by a continuous instrument Z or by the existence of independent covariates (Carneiro et al. 2011).

[9] Appendix A.6 contains a proof of this claim based on the Local Instrumental Variable (LIV) approach described by Heckman & Vytlacil (2005).

5 Partial identification when the support of the propensity score is discrete

When the support of the propensity score is not an interval, I cannot point-identify $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$, and $\Delta_S(x,u)$ without extra assumptions, implying that I cannot identify the bounds on $\Delta_{Y^*}^{OO}(x,u)$ given by corollary 10. There are two solutions for such lack of identification: I can non-parametrically bound those four objects (Mogstad et al. (2017)) or I can impose flexible parametric assumptions (Brinch et al. (2017)) to point-identify them. While the first approach is discussed in subsection 5.1, the second one is detailed in subsection 5.2.

5.1 Non-parametric outer set around the MTE of interest

For any $u \in [0,1]$ and $x \in \mathcal{X}$, I can bound $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$, and $\Delta_S(x,u)$ using the machinery proposed by Mogstad et al. (2017). To do so, fix $A \in \{Y,S\}$ and $d \in \{0,1\}$ and define the pair of functions $m^A \coloneqq \left(m_0^A, m_1^A\right)$ and the set of admissible MTR functions $\mathcal{M}^A \ni m^A$. Furthermore, fix $(x,u) \in \mathcal{X} \times [0,1]$ and define the functions $\Gamma_1^*\colon \mathcal{M}^Y \to \mathbb{R}$, $\Gamma_2^*\colon \mathcal{M}^Y \to \mathbb{R}$, $\Gamma_3^*\colon \mathcal{M}^S \to \mathbb{R}$ and $\Gamma_4^*\colon \mathcal{M}^S \to \mathbb{R}$ as:

$$\begin{aligned}
\Gamma_1^*\left(\tilde{m}^Y\right) &= \tilde{m}_1^Y(x,u) + 0 \cdot \tilde{m}_0^Y(x,u) \\
\Gamma_2^*\left(\tilde{m}^Y\right) &= 0 \cdot \tilde{m}_1^Y(x,u) + \tilde{m}_0^Y(x,u) \\
\Gamma_3^*\left(\tilde{m}^S\right) &= 0 \cdot \tilde{m}_1^S(x,u) + \tilde{m}_0^S(x,u) \\
\Gamma_4^*\left(\tilde{m}^S\right) &= \tilde{m}_1^S(x,u) - \tilde{m}_0^S(x,u),
\end{aligned}$$

and observe that $\Gamma_1^*\left(m^Y\right) = m_1^Y(x,u)$, $\Gamma_2^*\left(m^Y\right) = m_0^Y(x,u)$, $\Gamma_3^*\left(m^S\right) = m_0^S(x,u)$, and $\Gamma_4^*\left(m^S\right) = \Delta_S(x,u)$. Moreover, define, for each $A \in \{Y,S\}$, $\mathcal{G}_A$ as a collection of known or identified measurable functions $g_A\colon \{0,1\} \times \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ whose second moment is finite. For each IV-like specification $g_A \in \mathcal{G}_A$, define also $\beta_{g_A} \coloneqq \mathbb{E}\left[g_A(D,Z)\,A\,|X=x\right]$. According to proposition 1 by Mogstad et al. (2017), the function $\Gamma_{g_A}\colon \mathcal{M}^A \to \mathbb{R}$, defined as

$$\begin{aligned}\Gamma_{g_A}\left(\tilde{m}^A\right) = & \,\mathbb{E}\left[\int_0^1 \tilde{m}_0^A\left(X, u\right) \cdot g_A\left(0, Z\right) \cdot \mathbf{1}\left\{p\left(W\right) < u\right\} \mathrm{d}u \middle| X = x\right] \\ & + \mathbb{E}\left[\int_0^1 \tilde{m}_1^A\left(X, u\right) \cdot g_A\left(1, Z\right) \cdot \mathbf{1}\left\{p\left(W\right) \geq u\right\} \mathrm{d}u \middle| X = x\right],\end{aligned}$$

satisfies $\Gamma_{g_A}\left(m^A\right) = \beta_{g_A}$. As a result, m^A must lie in the set $\mathcal{M}_{\mathcal{G}_A}$ of admissible functions that satisfy the restrictions imposed by the data through the IV-like specifications, where:

$$\mathcal{M}_{\mathcal{G}_A} \coloneqq \left\{\tilde{m}^A \in \mathcal{M}^A\colon \Gamma_{g_A}\left(\tilde{m}^A\right) = \beta_{g_A} \text{ for all } g_A \in \mathcal{G}_A\right\}.$$

Assuming that $\mathcal{M}^A$ is convex and $M_{\mathcal{G}_A} \neq \emptyset$ for every $A \in \{Y, S\}$, proposition 2 by Mogstad et al. (2017) ensures that:

$$\begin{array}{rcccl}\inf\limits_{\tilde{m}^Y \in \mathcal{M}_{\mathcal{G}_Y}} \Gamma_1^*\left(\tilde{m}^Y\right) & \eqqcolon \underline{m_1^Y\left(x,u\right)} \leq & m_1^Y\left(x,u\right) & \leq \overline{m_1^Y\left(x,u\right)} \coloneqq & \sup\limits_{\tilde{m}^Y \in \mathcal{M}_{\mathcal{G}_Y}} \Gamma_3^*\left(\tilde{m}^Y\right) \\ \inf\limits_{\tilde{m}^Y \in \mathcal{M}_{\mathcal{G}_Y}} \Gamma_2^*\left(\tilde{m}^Y\right) & \eqqcolon \underline{m_0^Y\left(x,u\right)} \leq & m_0^Y\left(x,u\right) & \leq \overline{m_0^Y\left(x,u\right)} \coloneqq & \sup\limits_{\tilde{m}^Y \in \mathcal{M}_{\mathcal{G}_Y}} \Gamma_2^*\left(\tilde{m}^Y\right) \\ \inf\limits_{\tilde{m}^S \in \mathcal{M}_{\mathcal{G}_S}} \Gamma_3^*\left(\tilde{m}^S\right) & \eqqcolon \underline{m_0^S\left(x,u\right)} \leq & m_0^S\left(x,u\right) & \leq \overline{m_0^S\left(x,u\right)} \coloneqq & \sup\limits_{\tilde{m}^S \in \mathcal{M}_{\mathcal{G}_S}} \Gamma_3^*\left(\tilde{m}^S\right) \\ \inf\limits_{\tilde{m}^S \in \mathcal{M}_{\mathcal{G}_S}} \Gamma_4^*\left(\tilde{m}^S\right) & \eqqcolon \underline{\Delta_S\left(x,u\right)} \leq & \Delta_S\left(x,u\right) & \leq \overline{\Delta_S\left(x,u\right)} \coloneqq & \sup\limits_{\tilde{m}^S \in \mathcal{M}_{\mathcal{G}_S}} \Gamma_4^*\left(\tilde{m}^S\right)\end{array} \tag{24}$$

As a consequence, I can combine corollary 10 and inequalities (24) to provide a non-parametrically identified outer set around $\Delta_{Y^*}^{OO}\left(x,u\right)$:

Corollary 13 *Fix $u \in [0,1]$ and $x \in \mathcal{X}$ arbitrarily.*

Under assumptions 1-6, 7.1 and 8, the bounds of an outer set around $\Delta_{Y^*}^{OO}\left(x,u\right)$ are given by

$$\Delta_{Y^*}^{OO}\left(x,u\right) \geq \underline{y^*} - \frac{\overline{m_0^Y\left(x,u\right)}}{\underline{m_0^S\left(x,u\right)}}, \tag{25}$$

and that

$$\Delta_{Y^*}^{OO}\left(x,u\right) \leq \frac{\overline{m_1^Y\left(x,u\right)}}{\underline{m_0^S\left(x,u\right)}} - \frac{\underline{y^*} \cdot \underline{\Delta_S\left(x,u\right)}}{\overline{m_0^S\left(x,u\right)}} - \frac{\underline{m_0^Y\left(x,u\right)}}{\overline{m_0^S\left(x,u\right)}}. \tag{26}$$

Under assumptions 1-6, 7.2 and 8, the bounds of an outer set around $\Delta_{Y^*}^{OO}\left(x,u\right)$ are given *by*

$$\Delta_{Y^*}^{OO}\left(x,u\right) \geq \frac{\underline{m_1^Y\left(x,u\right)}}{\overline{m_0^S\left(x,u\right)}} - \frac{\overline{y}^* \cdot \overline{\Delta_S\left(x,u\right)}}{\underline{m_0^S\left(x,u\right)}} - \frac{\overline{m_0^Y\left(x,u\right)}}{\underline{m_0^S\left(x,u\right)}}, \tag{27}$$

and that

$$\Delta_{Y^*}^{OO}\left(x,u\right) \leq \overline{y}^* - \frac{\underline{m_0^Y\left(x,u\right)}}{\overline{m_0^S\left(x,u\right)}}. \tag{28}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)) and 8, the bounds of an outer set around $\Delta_{Y^*}^{OO}\left(x,u\right)$ are given by

$$\Delta_{Y^*}^{OO}\left(x,u\right) \geq \max\left\{\max\left\{\frac{\underline{m_1^Y\left(x,u\right)}}{\overline{m_0^S\left(x,u\right)}} - \frac{\overline{y}^* \cdot \overline{\Delta_S\left(x,u\right)}}{\underline{m_0^S\left(x,u\right)}}, \underline{y}^*\right\} - \frac{\overline{m_0^Y\left(x,u\right)}}{\underline{m_0^S\left(x,u\right)}}, \underline{y}^* - \overline{y}^*\right\}, \tag{29}$$

and that

$$\Delta_{Y^*}^{OO}\left(x,u\right) \leq \min\left\{\min\left\{\frac{\overline{m_1^Y\left(x,u\right)}}{\underline{m_0^S\left(x,u\right)}} - \frac{\underline{y}^* \cdot \underline{\Delta_S\left(x,u\right)}}{\overline{m_0^S\left(x,u\right)}}, \overline{y}^*\right\} - \frac{\underline{m_0^Y\left(x,u\right)}}{\overline{m_0^S\left(x,u\right)}}, \overline{y}^* - \underline{y}^*\right\}. \tag{30}$$

I stress that the cost of non-parametric partial identification of $m_0^Y\left(x,u\right)$, $m_1^Y\left(x,u\right)$, $m_0^S\left(x,u\right)$, and $\Delta_S\left(x,u\right)$ is losing the point-wise sharpness of the bounds around the target parameter $\Delta_{Y^*}^{OO}$. For that reason, corollary 13 is stated in terms of bounds of an outer set around $\Delta_{Y^*}^{OO}\left(x,u\right)$, that contains the true target parameter $\Delta_{Y^*}^{OO}\left(x,u\right)$ by construction.

5.2 Parametric identification of the MTE bounds

The fully non-parametric approach explained in subsection 5.1 may provide an uninformative outer set (i.e., equal to $\overline{y}^* - \underline{y}^*$ or $\underline{y}^* - \overline{y}^*$ under assumption 7.3). In such cases, parametric assumptions on the marginal treatment response function may buy a lot of identifying power. Although restrictive in principle, parametric assumptions may be flexible enough to provide credible bounds on $\Delta_{Y^*}^{OO}\left(x,u\right)$.

I fix $x \in \mathcal{X}$ and assume that the support of the propensity score $P\left(x,Z\right)$ is discrete and given by $\mathcal{P}_x = \{p_{x,1}, \ldots, p_{x,N}\}$ for some $N \in \mathbb{N}$. I could directly apply the identification strategy proposed by Brinch et al. (2017) by assuming that the MTR functions associated to Y and S are polynomial functions of U. However, this assumption is problematic for binary variables, such as the selection indicator S. For this reason, I make a small modification to the procedure suggested by Brinch et al. (2017): for $d \in \{0, 1\}$ and $A \in \{Y, S\}$, the MTR function is given by

$$m_d^A(x, u) = M^A\left(u, \boldsymbol{\theta}_{x,d}^A\right) \tag{31}$$

for any $u \in [0, 1]$, where $\Theta_x^A \subset \mathbb{R}^{2L}$ is a set of feasible parameters, $L \in \{0, \ldots, N-1\}$ is the number of parameters for each treatment group d, $\left(\boldsymbol{\theta}_{x,0}^A, \boldsymbol{\theta}_{x,1}^A\right) \in \Theta_x^A$ is a vector of pseudo-true unknown parameters, and $M^A \colon [0, 1] \times \mathbb{R}^{2L} \to \mathbb{R}$ is a known function. For example, in the case of a binary variable, a reasonable choice of M^A is the Bernstein Polynomial $\left(M^A\left(u, \boldsymbol{\theta}_{x,d}^A\right) = \sum_{l=0}^{L-1} \theta_{x,d,l}^A \cdot \binom{L}{l} \cdot u^l \cdot (1-u)^{L-l}\right)$ with feasible set $\Theta_x^A = [0, 1]^{2L}$. In the case of the selection indicator, the feasible set would be restricted by assumption 8 to $\Theta_x^A = \left\{\left(\tilde{\boldsymbol{\theta}}_{x,0}^A, \tilde{\boldsymbol{\theta}}_{x,1}^A\right) \in [0, 1]^{2L} : \tilde{\boldsymbol{\theta}}_{x,1}^A \geq \tilde{\boldsymbol{\theta}}_{x,0}^A\right\}$. I stress that the only difference between the Bernstein polynomial model and the simple polynomial model proposed by Brinch et al. (2017) is that it is easier to impose feasibility restrictions on the former model.

Back to the parametric model given by equation (31), I define the parameters $\left(\boldsymbol{\theta}_{x,0}^A, \boldsymbol{\theta}_{x,1}^A\right)$ as pseudo-true parameters in the sense that the parametric model in equation (31) is an approximation to the true data generating process via the moments $\mathbb{E}\left[A \,|X = x, P(W) = p_n, D = d\right]$ for any $d \in \{0, 1\}$ and $n \in \{1, \ldots, N\}$. Formally, I define

$$\begin{aligned}\left(\boldsymbol{\theta}_{x,0}^A, \boldsymbol{\theta}_{x,1}^A\right) := \underset{\left(\tilde{\boldsymbol{\theta}}_{x,0}^A, \tilde{\boldsymbol{\theta}}_{x,1}^A\right) \in \Theta_x^A}{\operatorname{argmin}} \sum_{n=1}^{N} \Bigg\{ & \left(\mathbb{E}\left[A \,|X = x, P(W) = p_n, D = 0\right] - \frac{\int_{p_n}^{1} M^A\left(u, \tilde{\boldsymbol{\theta}}_{x,0}^A\right) \mathrm{d}u}{1 - p_n}\right)^2 \\ & + \left(\mathbb{E}\left[A \,|X = x, P(W) = p_n, D = 1\right] - \frac{\int_{0}^{p_n} M^A\left(u, \tilde{\boldsymbol{\theta}}_{x,1}^A\right) \mathrm{d}u}{p_n}\right)^2 \Bigg\}.\end{aligned} \tag{32}$$

Note that, to estimate parameters $\left(\boldsymbol{\theta}_{x,0}^A, \boldsymbol{\theta}_{x,1}^A\right)$, I can simply use the sample analogue of equation (32), i.e., I only have to estimate a constrained OLS regression whose restrictions are given by the set Θ_x^A. The main advantage of estimating this parametric model using OLS is that I can easily test the model restrictions imposed through the set of feasible parameters

Θ_x^A. If such restrictions are valid and $L = N - 1$, then my parametric model collapses to the model proposed by Brinch et al. (2017) and I find that[10], for any $p_n \in \mathcal{P}_x$,

$$\mathbb{E}\left[A \,|X = x, P\left(W\right) = p_n, D = 0\right] = \frac{\int_{p_n}^{1} M^A\left(u, \boldsymbol{\theta}_{x,0}^A\right) \mathrm{d}u}{1 - p_n} \tag{33}$$

$$\mathbb{E}\left[A \,|X = x, P\left(W\right) = p_n, D = 1\right] = \frac{\int_{0}^{p_n} M^A\left(u, \boldsymbol{\theta}_{x,1}^A\right) \mathrm{d}u}{p_n}. \tag{34}$$

I can, then, combine corollary 10 and equation (31) and (32) to bound $\Delta_{Y^*}^{OO}(x, u)$:

Corollary 14 *Fix $u \in [0, 1]$ and $x \in \mathcal{X}$ arbitrarily.*

Under assumptions 1-6, 7.1 and 8, the bounds on $\Delta_{Y^*}^{OO}(x, u)$ are given by

$$\Delta_{Y^*}^{OO}(x, u) \geq \underline{y}^* - \frac{M^Y\left(u, \boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)}, \tag{35}$$

and that

$$\Delta_{Y^*}^{OO}(x, u) \leq \frac{M^Y\left(u, \boldsymbol{\theta}_{x,1}^Y\right) - \underline{y}^* \cdot \left[M^S\left(u, \boldsymbol{\theta}_{x,1}^S\right) - M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)\right]}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)} - \frac{M^Y\left(u, \boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)}. \tag{36}$$

Under assumptions 1-6, 7.2 and 8, the bounds on $\Delta_{Y^*}^{OO}(x, u)$ are given by

$$\Delta_{Y^*}^{OO}(x, u) \geq \frac{M^Y\left(u, \boldsymbol{\theta}_{x,1}^Y\right) - \overline{y}^* \cdot \left[M^S\left(u, \boldsymbol{\theta}_{x,1}^S\right) - M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)\right]}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)} - \frac{M^Y\left(u, \boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)}, \tag{37}$$

and that

$$\Delta_{Y^*}^{OO}(x, u) \leq \overline{y}^* - \frac{M^Y\left(u, \boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)}. \tag{38}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)) and 8, the bounds on $\Delta_{Y^*}^{OO}(x, u)$ are given by

$$\Delta_{Y^*}^{OO}(x, u) \geq \max\left\{\frac{M^Y\left(u, \boldsymbol{\theta}_{x,1}^Y\right) - \overline{y}^* \cdot \left[M^S\left(u, \boldsymbol{\theta}_{x,1}^S\right) - M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)\right]}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)}, \underline{y}^*\right\} - \frac{M^Y\left(u, \boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u, \boldsymbol{\theta}_{x,0}^S\right)}, \tag{39}$$

[10] Appendix A.7 contains a proof of this claim

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \min\left\{\frac{M^Y\left(u,\boldsymbol{\theta}_{x,1}^Y\right) - \underline{y}^* \cdot \left[M^S\left(u,\boldsymbol{\theta}_{x,1}^S\right) - M^S\left(u,\boldsymbol{\theta}_{x,0}^S\right)\right]}{M^S\left(u,\boldsymbol{\theta}_{x,0}^S\right)}, \overline{y}^*\right\} - \frac{M^Y\left(u,\boldsymbol{\theta}_{x,0}^Y\right)}{M^S\left(u,\boldsymbol{\theta}_{x,0}^S\right)}. \tag{40}$$

6 Further Work

This text is a working paper and still requires a few steps before it is finished. Currently, I am working on a empirical application in which I analyze the wage effect of the Job Corps Training Program. Such analysis builds upon the work by Lee (2009), Blanco et al. (2013), Chen & Flores (2015) and Chen et al. (2017) by looking at the entire MTE function and focusing on the heterogeneity of the treatment effect on wages. Another important future step is implementing a Monte Carlo experiment that can measure the informativeness of the proposed MTE bounds in a specific data generating process.

References

Ahn, H. & Powell, J. L. (1993), 'Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism', *Journal of Econometrics* **58**, pp. 3–29.

Altonji, J. (1993), 'The Demand for and Return to Education When Education Outcomes are Uncertain', *Journal of Labor Economics* **11**(1), pp. 48–83.

Angelucci, M., Karlan, D. & Zinman, J. (2015), 'Microcredit Impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Comportamos Banco', *American Economic Journal: Applied Economics* **7**(1), pp. 151–182.

Angrist, J., Bettinger, E. & Kremer, M. (2006), 'Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia', *The American Economic Review* **96**(3), 847–862.
URL: *http://www.jstor.org/stable/30034075*

Angrist, J., Lang, D. & Oreopoulos, P. (2009), 'Incentives and Services for College Achievement: Evidence from a Randomized Trial', *American Economic Journal: Applied Economics* **1**(1), pp. 1–28.

Behaghel, L., Crepon, B., Gurgand, M. & Barbanchon, T. L. (2015), 'Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models', *The Review of Economics and Statistics* **97**(5), pp. 1070–1080.

Bhattacharya, J., Shaikh, A. M. & Vytlacil, E. (2008), 'Treatment Effect Bounds under Monotonicity Assumptions: An Application to Swan-Ganz Catheterization', *The American Economic Review: Papers and Proceedings* **98**(2), pp. 351–356.

Bhattacharya, J., Shaikh, A. M. & Vytlacil, E. (2012), 'Treatment Effect Bounds: An Application to Swan-Ganz Catheterization', *Journal of Econometrics* **168**(2), pp. 223–243.

Blanco, G., Chen, X., Flores, C. A. & Flores-Lagunes, A. (2017), Bounding Average and Quantile Effects of Training on Employment and Unemployment Durations under Selection, Censoring, and Noncompliance. Available at: http://conference.iza.org/conference_files/EVAL_2017/blanco_g7367.pdf.

Blanco, G., Flores, C. A. & Flores-Lagunes, A. (2013), 'Bounds on Average and Quantile Treatment Effects of Job Corps Training on Wages', *Journal of Human Resources* **48**(3), pp. 659–701.

Blundell, R., Gosling, A., Ichimura, H. & Meghir, C. (2007), 'Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds', *Econometrica* **75**(2), pp. 323–363.

Brinch, C. N., Mogstad, M. & Wiswall, M. (2017), 'Beyond LATE with a Discrete Instrument', *Journal of Political Economy* **125**(4), pp. 985–1039.

Card, D. (1999), The Causal Effect of Education on Earnings, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3A, Elsevier, pp. pp. 1801–1863.

Carneiro, P., Heckman, J. J. & Vytlacil, E. (2011), 'Estimating Marginal Returns to Education', *The American Economic Review* **101**(6), pp. 2754–2781.

CASS (1984), 'Myocardial Infarction and Mortality in the Coronary Artery Surgery Study (CASS) Randomized Trial', *The New England Journal of Medicine* **310**(12), pp. 750–758.

Chen, X. & Flores, C. A. (2015), 'Bounds on Treatment Effects in the Presence of Sample Selection and Noncompliance: The Wage Effects of Job Corps', *Journal of Business and Economic Statistics* **33**(4), pp. 523–540.

Chen, X., Flores, C. A. & Flores-Lagunes, A. (2017), 'Going beyond LATE: Bounding Average Treatment Effects of Job Corps Training', *Journal of Human Resources* .

Chen, X., Hong, H. & Tarozzi, A. (2008), 'Semiparametric Efficiency in GMM Models with Auxiliary Data', *The Annals of Statistics* **36**(2), 808–843.

Chesher, A. (2010), 'Instrumental Variable Models for Discrete Outcomes', *Econometrica* **78**(2), pp. 575–601.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W. & Yagan, D. (2011), 'How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star', *The Quarterly Journal of Economics* **126**(4), 1593–1660.
URL: *http://dx.doi.org/10.1093/qje/qjr041*

Chiburis, R. C. (2010), 'Semiparametric Bounds on Treatment Effects', *Journal of Econometrics* **159**(2), 267–275.

DeMel, S., McKenzie, D. & Woodruff, C. (2013), 'The Demand for, and Consequences of, Formalization among Informal First in Sri Lanka', *American Economic Journal: Applied Economics* **5**(2), 122–150.

Dobbie, W. & Jr., R. G. F. (2015), 'The Medium-Term Impacts of High-Achieving Charter Schools', *Journal of Political Economy* **123**(5), pp. 985–1037.

Farber, H. S. (1993), 'The Incidence and Costs of Job Loss: 1982-91', *Brookings Papers: Microeconomics* pp. pp. 73–132.

Frangakis, C. & Rubin, D. (2002), 'Principal Stratification in Causal Inference', *Biometrics* **58**(1), pp. 21–29.

Fricke, H., Frolich, M., Huber, M. & Lechner, M. (2015), Endogeneity and Non-Response Bias in Treatment Evaluation: Nonparametric Identification of Causal Effects by Instruments. Available at: https://www.iza.org/publications/dp/9428/endogeneity-and-non-response-bias-in-treatment-evaluation-nonparametric-identification-of-

Frolich, M. & Huber, M. (2014), 'Treatment Evaluation with Multiple Outcome Periods Under Endogeneity and Attrition', *Journal of the American Statistical Association* **109**(508), 1697–1711.

Heckman, J. J. (1979), 'Sample Selection Bias as a Specification Error', *Econometrica* **47**(1), pp. 153–161.

Heckman, J. J. & Borjas, G. J. (1980), 'Does unemployment cause future unemployment? definitions, questions and answers from a continuous time model of heterogeneity and state dependence', *Economica* **47**(187), 247–283.
URL: *http://www.jstor.org/stable/2553150*

Heckman, J. J., Urzua, S. & Vytlacil, E. (2006), 'Understanding Instrumental Variables in Models with Essential Heterogeneity', *The Review of Economics and Statistics* **88**(3), pp. 389–432.

Heckman, J. J. & Vytlacil, E. (2001*a*), Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect, *in* M. Lechner & J. Pfeifer, eds, 'Econometric Evaluation of Labour Market Policies', Vol. 13 of *ZEW Economic Studies*, Physica, Heidelberg.

Heckman, J. J. & Vytlacil, E. (2005), 'Structural Equations, Treatment Effects and Econometric Policy Evaluation', *Econometrica* **73**(3), pp. 669–738.

Heckman, J., LaLonde, R. & Smith, J. (1999), The Economics and Econometrics of Active Labor Market Programs, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3A, Elsevier, pp. pp. 1865–2097.

Heckman, J. & Vytlacil, E. (1999), 'Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects', *Proceedings of the National Academy of Sciences of the United States of America* **96**, 4730–4734.

Heckman, J. & Vytlacil, E. (2001*b*), 'Policy-Relevant Treatment Effects', *American Economic Review: Papers and Proceedings* **91**(2), pp. 107–111.

Helland, E. & Yoon, J. (2017), 'Estimating the Effects of the English Rule on Litigation Outcomes', *The Review of Economics and Statistics* **99**(4), pp. 678–682.

Huber, M. (2014), 'Treatment Evaluation in the Presence of Sample Selection', *Econometric Reviews* **33**(8), pp. 869–905.

Huber, M., Laffers, L. & Mellace, G. (2017), 'Sharp IV Bounds on Average Treatment Effects on the Treated and Other Populations under Endogeneity and Noncompliance', *Journal of Applied Econometrics* **32**, pp. 56–79.

Huber, M. & Mellace, G. (2015), 'Sharp Bounds on Causal Effects under Sample Selection', *Oxford Bulletin of Economics and Statistics* **77**(1), pp. 129–151.

Imai, K. (2008), 'Sharp Bounds on the Causal Effects in Randomized Experiments with Truncation- by- Death', *Statistics and Probability Letters* **78**(2), pp. 144–149.

Imbens, G. W. & Angrist, J. D. (1994), 'Identification and Estimation of Local Average Treatment Effects', *Econometrica* **62**(2), pp. 467–475.

Jacobson, L. S., LaLonde, R. J. & Sullivan, D. G. (1993), 'Earnings Losses of Displaced Workers', *The American Economic Review* **83**(4), pp. 685–709.

Krueger, A. B. & Whitmore, D. M. (2001), 'The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR', *The Economic Journal* **111**(468), pp. 1–28.

Lechner, M. & Mell, B. (2010), Partial Idendification of Wage Effects of Training Programs. Available at: https://ideas.repec.org/p/bro/econwp/2010-8.html.

Lee, D. S. (2009), 'Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects', *The Review of Economic Studies* **76**, pp. 1071–1102.

Lee, S. & Salanie, B. (2016), Identifying Effects of Multivalued Treatments.

Machado, C., Shaikh, A. M. & Vytlacil, E. (2018), Instrumental Variables and the Sign of the Average Treatment Effect. Forthcoming on the Journal of Econometrics. Available at: https://drive.google.com/file/d/1yIdbyzVaCj5f9cOHN8K_nhSKGDKDhCzY/view.

Manski, C. F. (1990), 'Nonparametric Bounds on Treatment Effects', *American Economic Review* **80**(2), pp. 319–323.

Manski, C. F. (1997), 'Monotone Treatment response', *Econometrica* **65**(6), pp. 1311–1334.

Manski, C. F. & Pepper, J. V. (2000), 'Monotone Instrumental Variables: With an Application to the Returns to Schooling', *Econometrica* **68**(4), pp. 997–1010.

Mealli, F. & Pacini, B. (2013), 'Using Secondary Outcomes to Sharpen Inference in Randomized Experiments with Noncompliance', *Journal of the American Statistical Association* **108**(503), pp. 1120–1131.

Mogstad, M., Santos, A. & Torgovitsky, A. (2017), Using Instrumental Variables for Inference about Policy Relevant Treatment Effects. NBER Working Paper 23568.

Newey, W. K., Powell, J. L. & Vella, F. (1999), 'Nonparametric Estimation of Triangular Simultaneous Equations Models', *Econometrica* **67**(3), pp. 565–603.

Rubin, D. (1974), 'Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies', *Journal of Educational Psychology* **66**(5), 688–701.

Shaikh, A. & Vytlacil, E. (2011), 'Partial Identification in Triangular Systems of Equations with Binary Dependent Variables', *Econometrica* **79**(3), pp. 949–955.

Steinmayr, A. (2014), When a Random Sample is not Random: Bounds on the Effect of Migration on Children Left Behind. Working Paper.

U.S. Department of Health and Human Services (2004), *The Health Consequences of Smoking: A Report of the Surgeon General*, U.S. Department of Health and Human Services, Public Health Service, Office on Smoking and Health. Available at: https://www.cdc.gov/tobacco/data_statistics/sgr/2004/index.htm.

Vytlacil, E. (2002), 'Independence, Monotonicity and Latent Index Models: An Equivalence Result', *Econometrica* **70**(1), pp. 331–341.

Supporting Information (Online Appendix)

A Proofs of the main results

A.1 Proof of Equation (4)

Note that

$$\begin{aligned}
\mathbb{E}\left[Y_0^* \,|X = x, U = u, S_0 = 1, S_1 = 1\right] &= \mathbb{E}\left[Y_0^* \,|X = x, U = u, S_0 = 1\right] \\
&\quad \text{by assumption 8} \\
&= \frac{\mathbb{E}\left[S_0 \cdot Y_0^* \,|X = x, U = u\right]}{\mathbb{P}\left[S_0 = 1 \,|X = x, U = u\right]} \\
&\quad \text{by the definition of conditional expectation} \\
&= \frac{\mathbb{E}\left[Y_0 \,|X = x, U = u\right]}{\mathbb{E}\left[S_0 \,|X = x, U = u\right]} \\
&= \frac{m_0^Y\left(x, u\right)}{m_0^S\left(x, u\right)}. \qquad \blacksquare
\end{aligned}$$

A.2 Proof of Equation (5)

First, observe that

$$\begin{aligned}
m_0^S\left(x, u\right) &\coloneqq \mathbb{E}\left[S_0 \,|X = x, U = u\right] \\
&= \mathbb{P}\left[Q\left(0, X\right) \geq V \,|X = x, U = u\right] \qquad \text{(A.1)} \\
&\quad \text{by equation (2),} \\
m_1^S\left(x, u\right) &\coloneqq \mathbb{E}\left[S_1 \,|X = x, U = u\right] \\
&= \mathbb{P}\left[Q\left(1, X\right) \geq V \,|X = x, U = u\right] \qquad \text{(A.2)} \\
&\quad \text{by equation (2),} \\
\Delta_S\left(x, u\right) &\coloneqq \mathbb{E}\left[S_1 - S_0 \,|X = x, U = u\right] \\
&= m_1^S\left(x, u\right) - m_0^S\left(x, u\right)
\end{aligned}$$

$$= \mathbb{P}\left[Q\left(1,X\right) \geq V > Q\left(0,X\right) \left| X = x, U = u\right.\right]$$

by equations (A.1) and (A.2) and assumption (8)

$$= \mathbb{P}\left[S_0 = 0, S_1 = 1 \left| X = x, U = u\right.\right] \tag{A.3}$$

by equation (2), and

$$\Delta_Y^{NO}\left(x,u\right) \coloneqq \mathbb{E}\left[Y_1 - Y_0 \left| X = x, U = u, S_0 = 0, S_1 = 1\right.\right]$$

$$= \mathbb{E}\left[S_1 \cdot Y_1^* - S_0 \cdot Y_0^* \left| X = x, U = u, S_0 = 0, S_1 = 1\right.\right]$$

$$= \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 0, S_1 = 1\right.\right]. \tag{A.4}$$

Note also that:

$$m_1^Y\left(x,u\right) \coloneqq \mathbb{E}\left[Y_1 \left| X = x, U = u\right.\right]$$

$$= \mathbb{E}\left[S_1 \cdot Y_1^* \left| X = x, U = u\right.\right]$$

$$= \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \cdot \mathbb{P}\left[S_0 = 1 \left| X = x, U = u\right.\right]$$

$$+ \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 0, S_1 = 1\right.\right] \cdot \mathbb{P}\left[S_0 = 0, S_1 = 1 \left| X = x, U = u\right.\right]$$

by assumption 8 and the Law of Iterated Expectations

$$= \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \cdot m_0^S\left(x,u\right) + \Delta_Y^{NO}\left(x,u\right) \cdot \Delta_S\left(x,u\right) \tag{A.5}$$

by equations (A.1), (A.3) and (A.4),

implying equation (5) after some rearrangement. ■

A.3 Proof of Proposition 9

Note that

$$\underline{y}^* \leq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \leq \overline{y}^* \tag{A.6}$$

by the definition of $\underline{y}^*$ and $\overline{y}^*$. Observe also that

$$\underline{y}^* \leq \Delta_Y^{NO}\left(x,u\right) \leq \overline{y}^*$$

by equation (A.4) and the definition of $\underline{y}^*$ and $\overline{y}^*$, implying, by equation (5), that

$$\mathbb{E}\left[Y_1^* \left| X=x, U=u, S_0=1, S_1=1\right.\right] \leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} \tag{A.7}$$

under assumption 7.1,

$$\frac{m_1^Y(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} \leq \mathbb{E}\left[Y_1^* \left| X=x, U=u, S_0=1, S_1=1\right.\right] \tag{A.8}$$

under assumption 7.2, and

$$\begin{aligned}\frac{m_1^Y(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} &\leq \mathbb{E}\left[Y_1^* \left| X=x, U=u, S_0=1, S_1=1\right.\right] \\ &\leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)}.\end{aligned} \tag{A.9}$$

under assumption 7.3 (sub-case (a) or (b)). Combining equations (A.6)-(A.9), it is easy to show that proposition 9 holds. ■

A.4 Proof of Proposition 11

First, I prove proposition 11 under assumption 7.3 (sub-cases (a) and (b)). At the end of this subsection, I prove proposition 11 under assumptions 7.1 and 7.2.

A.4.1 Proof under Assumption 7.3 (sub-cases (a) and (b))

Fix $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u})\right)$ arbitrarily. For brevity, define $\alpha(\overline{x},\overline{u}) := \delta(\overline{x},\overline{u}) + \dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})}$ and $\gamma(\overline{x},\overline{u}) := \dfrac{m_1^Y(\overline{x},\overline{u}) - \alpha(\overline{x},\overline{u}) \cdot m_0^S(\overline{x},\overline{u})}{\Delta_S(\overline{x},\overline{u})}$.

Note that

$$
\begin{aligned}
\delta\left(\overline{x},\overline{u}\right) &\in \left(\underline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right),\overline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right)\right) \\
\Leftrightarrow \quad \alpha\left(\overline{x},\overline{u}\right) &\in \left(\max\left\{\frac{m_1^Y\left(x,u\right)-\overline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)},\underline{y}^*\right\},\right. \\
&\qquad\qquad \left.\min\left\{\frac{m_1^Y\left(x,u\right)-\underline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)},\overline{y}^*\right\}\right) \\
&\subseteq \left(\underline{y}^*,\overline{y}^*\right),
\end{aligned}
\tag{A.10}
$$

and that

$$
\begin{aligned}
\alpha\left(\overline{x},\overline{u}\right) &\in \left(\frac{m_1^Y\left(x,u\right)-\overline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)},\frac{m_1^Y\left(x,u\right)-\underline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}\right) \\
\Leftrightarrow \quad \gamma\left(\overline{x},\overline{u}\right) &\in \left(\underline{y}^*,\overline{y}^*\right).
\end{aligned}
\tag{A.11}
$$

The strategy of this proof consists of defining random variables $\left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}\right)$ through their joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ and, then, checking that equations (15), (16) and (17) are satisfied. I fix $(y_0,y_1,u,v,z,x)\in\mathbb{R}^6$ and define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ in twelve steps:

Step 1. For $x\notin\mathcal{X}$, $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0,y_1,u,v,z,x\right)=F_{Y_0^*,Y_1^*,U,V,Z,X}\left(y_0,y_1,u,v,z,x\right)$.

Step 2. From now on, assume that $x\in\mathcal{X}$. Since

$$
F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0,y_1,u,v,z,x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\left|x\right.\right)\cdot F_X\left(x\right),
$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\left|x\right.\right)$. Moreover, I impose

$$
Z\perp\!\!\!\perp\left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}\right)|X
$$

by writing

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\,|x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)\cdot F_{Z|X}\left(z\,|x\right),$$

implying that it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 3. For $u\notin[0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)=F_{Y_0^*,Y_1^*,U,V|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 4. From now on, assume that $u\in[0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)\cdot F_{\tilde{U}|X}\left(u\,|x\right),$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)$ and $F_{\tilde{U}|X}\left(u\,|x\right)$.

Step 5. I define $F_{\tilde{U}|X}\left(u\,|x\right)=F_{U|X}\left(u\,|x\right)=u$.

Step 6. For any $u\neq\overline{u}$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)=F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,u\right)$.

Step 7. For any $v\notin[0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,\overline{u}\right)=F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,\overline{u}\right)$.

Step 8. From now on, assume that $v\in[0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,\overline{u}\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)\cdot F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right),$$

it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)$ and $F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)$.

Step 9. I define

$$F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)=\begin{cases} m_0^S\left(x,\overline{u}\right)\cdot\dfrac{v}{Q\left(0,x\right)} & \text{if } v\leq Q\left(0,x\right) \\ m_0^S\left(x,\overline{u}\right)+\Delta_S\left(x,\overline{u}\right)\cdot\dfrac{v-Q\left(0,x\right)}{Q\left(1,x\right)-Q\left(0,x\right)} & \text{if } Q\left(0,x\right)<v\leq Q\left(1,x\right) \\ m_1^S\left(x,\overline{u}\right)+\left(1-m_1^S\left(x,\overline{u}\right)\right)\dfrac{v-Q\left(1,x\right)}{1-Q\left(1,x\right)} & \text{if } Q\left(1,x\right)<v \end{cases}.$$

Step 10. I write $F_{\tilde{Y}_0^*, \tilde{Y}_1^* | X, \tilde{U}, \tilde{V}} \left(y_0, y_1 \left| x, \overline{u}, v \right. \right) = F_{\tilde{Y}_0^* | X, \tilde{U}, \tilde{V}} \left(y_0 \left| x, \overline{u}, v \right. \right) \cdot F_{\tilde{Y}_1^* | X, \tilde{U}, \tilde{V}} \left(y_1 \left| x, \overline{u}, v \right. \right)$, implying that I can separately define $F_{\tilde{Y}_0^* | X, \tilde{U}, \tilde{V}} \left(y_0 \left| x, \overline{u}, v \right. \right)$ and $F_{\tilde{Y}_1^* | X, \tilde{U}, \tilde{V}} \left(y_1 \left| x, \overline{u}, v \right. \right)$.

Step 11. When $\mathcal{Y}^*$ is a bounded interval (sub-case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_0^* | X, \tilde{U}, \tilde{V}} \left(y_0 \left| x, \overline{u}, v \right. \right) = \begin{cases} \mathbf{1} \left\{ y_0 \geq \dfrac{m_0^Y \left(\overline{x}, \overline{u} \right)}{m_0^S \left(\overline{x}, \overline{u} \right)} \right\} & \text{if } v \leq Q \left(0, x \right) \\ ------------ & -------- \\ \mathbf{1} \left\{ y_0 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2} \right\} & \text{if } Q \left(0, x \right) < v \end{cases}.$$

When $\overline{y}^* = \max \left\{ y \in \mathcal{Y}^* \right\}$ and $\underline{y}^* = \min \left\{ y \in \mathcal{Y}^* \right\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_0^* | X, \tilde{U}, \tilde{V}} \left(y_0 \left| x, \overline{u}, v \right. \right) = \begin{cases} 0 & \text{if } y_0 < \underline{y}^* \text{ and } v \leq Q \left(0, x \right) \\ 1 - \dfrac{\dfrac{m_0^Y \left(\overline{x}, \overline{u} \right)}{m_0^S \left(\overline{x}, \overline{u} \right)} - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_0 < \overline{y}^* \text{ and } v \leq Q \left(0, x \right) \\ 1 & \text{if } \overline{y}^* \leq y_0 \text{ and } v \leq Q \left(0, x \right) \\ ------------ & ----------------- \\ \mathbf{1} \left\{ y_0 \geq \overline{y}^* \right\} & \text{if } Q \left(0, x \right) < v \end{cases}.$$

which are valid cumulative distribution functions because $\dfrac{m_0^Y \left(\overline{x}, \overline{u} \right)}{m_0^S \left(\overline{x}, \overline{u} \right)} \in \left[\underline{y}^*, \overline{y}^* \right]$.

Step 12. When $\mathcal{Y}^*$ is a bounded interval (case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_1^* | X, \tilde{U}, \tilde{V}} \left(y_1 \left| x, \overline{u}, v \right. \right) = \begin{cases} \mathbf{1} \left\{ y_1 \geq \alpha \left(\overline{x}, \overline{u} \right) \right\} & \text{if } v \leq Q \left(0, x \right) \\ ---------- & ------------- \\ \mathbf{1} \left\{ y_1 \geq \gamma \left(\overline{x}, \overline{u} \right) \right\} & \text{if } Q \left(0, x \right) < v \leq Q \left(1, x \right) \\ ---------- & ------------- \\ \mathbf{1} \left\{ y_1 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2} \right\} & \text{if } Q \left(1, x \right) < v \end{cases}.$$

When $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_1|x,\overline{u},v) = \begin{cases} 0 & \text{if } y_1 < \underline{y}^* \text{ and } v \leq Q(0,x) \\ 1 - \dfrac{\alpha(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } v \leq Q(0,x) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } v \leq Q(0,x) \\ -------- & ------------------ \\ 0 & \text{if } y_1 < \underline{y}^* \text{ and } Q(0,x) < v \leq Q(1,x) \\ 1 - \dfrac{\gamma(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } Q(0,x) < v \leq Q(1,x) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } Q(0,x) < v \leq Q(1,x) \\ -------- & ------------------ \\ \mathbf{1}\{y_1 \geq \overline{y}^*\} & \text{if } Q(1,x) < v \end{cases}.$$

which are valid cumulative distribution functions because of equations (A.10) and (A.11).

Having defined the joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$, note that equations (A.10) and (A.11), $\dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} \in \left[\underline{y}^*, \overline{y}^*\right]$ and steps 7-12 ensure that equation (16) holds. Now, I show, in three steps, that equation (15) holds.

Step 13. Observe that

$$\begin{aligned} &\mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] \\ &\quad = \mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, Q(0,\overline{x}) \geq \tilde{V}\right] \\ &\qquad \text{by the definition of } \tilde{S}_0 \text{ and } \tilde{S}_1 \\ &\quad = \frac{\mathbb{E}\left[\mathbf{1}\left\{Q(0,\overline{x}) \geq \tilde{V}\right\} \cdot \tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]}{\mathbb{P}\left[Q(0,\overline{x}) \geq \tilde{V} \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]} \end{aligned}$$

by the definition of conditional expectation

$$= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q\left(0,\overline{x}\right) \geq \tilde{V}\right\} \cdot \mathbb{E}\left[\tilde{Y}_1^* \left| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V}\right.\right] \left| X = \overline{x}, \tilde{U} = \overline{u}\right.\right]}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \left| X = \overline{x}, \tilde{U} = \overline{u}\right.\right]}$$

by the law of iterated expectations

$$= \frac{\int\limits_0^{Q(0,\overline{x})} \mathbb{E}\left[\tilde{Y}_1^* \left| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V} = v\right.\right] \mathrm{d}F_{\tilde{V}|X,\tilde{U}}\left(v \left| x, \overline{u}\right.\right)}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \left| X = \overline{x}, \tilde{U} = \overline{u}\right.\right]}$$

by the definition of expectation and by step 7

$$= \frac{\int\limits_0^{Q(0,\overline{x})} \alpha\left(\overline{x},\overline{u}\right) \mathrm{d}F_{\tilde{V}|X,\tilde{U}}\left(v \left| x, \overline{u}\right.\right)}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \left| X = \overline{x}, \tilde{U} = \overline{u}\right.\right]}$$

by step 12

$$= \alpha\left(\overline{x},\overline{u}\right) \tag{A.12}$$

by linearity of the Lebesgue Integral

Step 14. Notice that

$$\mathbb{E}\left[\tilde{Y}_0^* \left| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right.\right]$$

$$= \mathbb{E}\left[\tilde{Y}_0^* \left| X = \overline{x}, \tilde{U} = \overline{u}, Q\left(0,\overline{x}\right) \geq \tilde{V}\right.\right]$$

by the definition of $\tilde{S}_0$ and $\tilde{S}_1$

$$= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q\left(0,\overline{x}\right) \geq \tilde{V}\right\} \cdot \tilde{Y}_0^* \left| X = \overline{x}, \tilde{U} = \overline{u}\right.\right]}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \left| X = \overline{x}, \tilde{U} = \overline{u}\right.\right]}$$

by the definition of conditional expectation

$$= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q\left(0,\overline{x}\right) \geq \tilde{V}\right\} \cdot \mathbb{E}\left[\tilde{Y}_0^* \left| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V}\right.\right] \left| X = \overline{x}, \tilde{U} = \overline{u}\right.\right]}{\mathbb{P}\left[Q\left(0,\overline{x}\right) \geq \tilde{V} \left| X = \overline{x}, \tilde{U} = \overline{u}\right.\right]}$$

by the law of iterated expectations

$$= \frac{\int\limits_{0}^{Q(0,\overline{x})} \mathbb{E}\left[\tilde{Y}_0^* \left| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V} = v \right.\right] \mathrm{d}F_{\tilde{V}|X,\tilde{U}}\left(v \left| x, \overline{u}\right.\right)}{\mathbb{P}\left[Q\left(0, \overline{x}\right) \geq \tilde{V} \left| X = \overline{x}, \tilde{U} = \overline{u} \right.\right]}$$

$$= \frac{\int\limits_{0}^{Q(0,\overline{x})} \frac{m_0^Y\left(\overline{x}, \overline{u}\right)}{m_0^S\left(\overline{x}, \overline{u}\right)} \mathrm{d}F_{\tilde{V}|X,\tilde{U}}\left(v \left| x, \overline{u}\right.\right)}{\mathbb{P}\left[Q\left(0, \overline{x}\right) \geq \tilde{V} \left| X = \overline{x}, \tilde{U} = \overline{u} \right.\right]}$$

by step 11

$$= \frac{m_0^Y\left(\overline{x}, \overline{u}\right)}{m_0^S\left(\overline{x}, \overline{u}\right)}. \tag{A.13}$$

Step 15. Note that

$$\begin{aligned}
\Delta_{\tilde{Y}^*}^{OO}\left(\overline{x}, \overline{u}\right) &\coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \left| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1 \right.\right] \\
&= \mathbb{E}\left[\tilde{Y}_1^* \left| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1 \right.\right] \\
&\quad - \mathbb{E}\left[\tilde{Y}_0^* \left| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1 \right.\right] \\
&= \alpha\left(\overline{x}, \overline{u}\right) - \frac{m_0^Y\left(\overline{x}, \overline{u}\right)}{m_0^S\left(\overline{x}, \overline{u}\right)} \\
&\qquad \text{by equations (A.12) and (A.13)} \\
&= \delta\left(\overline{x}, \overline{u}\right) \\
&\qquad \text{by the definition of } \alpha\left(\overline{x}, \overline{u}\right),
\end{aligned}$$

ensuring that equation (15) holds.

Finally, I show, in two steps, that equation (17) holds.

Step 16. Fix $(y, d, s, z) \in \mathbb{R}^4$ arbitrarily and observe that equation (17) can be simplified to:

$$\begin{aligned}
&F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}\left(y, d, s, z, \overline{x}\right) = F_{Y,D,S,Z,X}\left(y, d, s, z, \overline{x}\right) \\
\Leftrightarrow &F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}\left(y, d, s, z \left| \overline{x}\right.\right) \cdot F_X\left(\overline{x}\right) = F_{Y,D,S,Z|X}\left(y, d, s, z \left| \overline{x}\right.\right) \cdot F_X\left(\overline{x}\right) \\
\Leftrightarrow &F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}\left(y, d, s, z \left| \overline{x}\right.\right) = F_{Y,D,S,Z|X}\left(y, d, s, z \left| \overline{x}\right.\right)
\end{aligned} \tag{A.14}$$

Step 17. Notice that

$$
\begin{aligned}
& F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}\left(y,d,s,z\left|\overline{x}\right.\right) \\
&\quad = \mathbb{E}\left[\mathbf{1}\left\{\left(\tilde{Y},\tilde{D},\tilde{S},Z\right)\leq(y,d,s,z)\right\}\middle| X=\overline{x}\right] \\
&\quad = \int \mathbf{1}\left\{\left(\tilde{Y},\tilde{D},\tilde{S},Z\right)\leq(y,d,s,z)\right\}\mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\left|\overline{x}\right.\right) \\
&\qquad \text{because } \left(\tilde{Y},\tilde{D},\tilde{S},Z\right) \text{ are functions of } \left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z\right) \\
&\quad = \int \left[\mathbf{1}\left\{\left(\tilde{Y},\tilde{D},\tilde{S},Z\right)\leq(y,d,s,z)\right\}\cdot\mathbf{1}\left\{u\neq\overline{u}\right\}\right]\mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\left|\overline{x}\right.\right) \\
&\qquad + \int \left[\mathbf{1}\left\{\left(\tilde{Y},\tilde{D},\tilde{S},Z\right)\leq(y,d,s,z)\right\}\cdot\mathbf{1}\left\{u=\overline{u}\right\}\right]\mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\left|\overline{x}\right.\right) \\
&\qquad \text{by linearity of the Lebesgue Integral} \\
&\quad = \int \left[\mathbf{1}\left\{\left(\tilde{Y},\tilde{D},\tilde{S},Z\right)\leq(y,d,s,z)\right\}\cdot\mathbf{1}\left\{u\neq\overline{u}\right\}\right]\mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\left|\overline{x}\right.\right) \\
&\qquad \text{because } \mathbb{P}\left[\tilde{U}=\overline{u}\middle| X=\overline{x}\right]=0 \text{ by step 5} \\
&\quad = \int \left[\mathbf{1}\left\{(Y,D,S,Z)\leq(y,d,s,z)\right\}\cdot\mathbf{1}\left\{u\neq\overline{u}\right\}\right]\mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0,y_1,u,v,z\left|\overline{x}\right.\right) \\
&\qquad \text{by steps 2-6} \\
&\quad = \int \left[\mathbf{1}\left\{(Y,D,S,Z)\leq(y,d,s,z)\right\}\cdot\mathbf{1}\left\{u\neq\overline{u}\right\}\right]\mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0,y_1,u,v,z\left|\overline{x}\right.\right) \\
&\qquad + \int \left[\mathbf{1}\left\{(Y,D,S,Z)\leq(y,d,s,z)\right\}\cdot\mathbf{1}\left\{u=\overline{u}\right\}\right]\mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0,y_1,u,v,z\left|\overline{x}\right.\right) \\
&\qquad \text{because } \mathbb{P}\left[U=\overline{u}\left|X=\overline{x}\right.\right]=0 \\
&\quad = \int \mathbf{1}\left\{(Y,D,S,Z)\leq(y,d,s,z)\right\}\mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0,y_1,u,v,z\left|\overline{x}\right.\right) \\
&\qquad \text{by linearity of the Lebesgue Integral} \\
&\quad = \mathbb{E}\left[\mathbf{1}\left\{(Y,D,S,Z)\leq(y,d,s,z)\right\}\middle| X=\overline{x}\right] \\
&\quad = F_{Y,D,S,Z|X}\left(y,d,s,z\left|\overline{x}\right.\right),
\end{aligned}
$$

implying equation (17) according to equation (A.14).

I can, then, conclude that proposition 11 is true. ■

As a remark, the above constructive proof defines random variables $\left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}\right)$ that matches other important moments of the true data generating process besides the ones imposed by proposition 11.

Remark 1. Note that

$$\mathbb{P}\left[\tilde{S}_0 = 1, \tilde{S}_1 = 1 \middle| X = \overline{x}, \tilde{U} = \overline{u}\right] = \mathbb{P}\left[Q(0, \overline{x}) \geq \tilde{V} \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]$$

$$\text{by the definition of } \tilde{S}_0 \text{ and } \tilde{S}_1$$

$$= m_0^S(\overline{x}, \overline{u}) \tag{A.15}$$

$$\text{by step 9.}$$

and that

$$\mathbb{P}\left[\tilde{S}_0 = 0, \tilde{S}_1 = 1 \middle| X = \overline{x}, \tilde{U} = \overline{u}\right] = \mathbb{P}\left[Q(1, \overline{x}) \geq \tilde{V} > Q(0, \overline{x}) \middle| X = \overline{x}, \tilde{U} = \overline{u}\right]$$

$$\text{by the definition of } \tilde{S}_0 \text{ and } \tilde{S}_1$$

$$= \Delta_S(\overline{x}, \overline{u}) \tag{A.16}$$

$$\text{by step 9.}$$

Remark 2. Analogously to equation (A.12), I find that

$$\mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 0, \tilde{S}_1 = 1\right] = \gamma(\overline{x}, \overline{u}). \tag{A.17}$$

Remark 3. Combining equations (A.5) and (A.15)-(A.17), I have that

$$\mathbb{E}\left[\tilde{Y}_1 \middle| X = x, \tilde{U} = \overline{u}\right] = m_1^Y(\overline{x}, \overline{u}).$$

A.4.2 Proof under Assumptions 7.1 and 7.2

I, now, prove proposition 11 under assumptions 7.1 and 7.2. In particular, I focus on the case $\underline{y}^* > -\infty$ and $\overline{y}^* = +\infty$ (assumption 7.1) because it is more common in empirical applications. The case $\underline{y}^* = -\infty$ and $\overline{y}^* < +\infty$ (assumption 7.2) is symmetric.

The proof under assumption 7.1 is equal to the proof under assumption 7.3(a). The only difference is that

$$\begin{aligned}
&\delta\left(\overline{x},\overline{u}\right) &&\in \left(\underline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right),\overline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right)\right)\\
\Leftrightarrow\ &\alpha\left(\overline{x},\overline{u}\right) &&\in \left(\underline{y}^*,\frac{m_1^Y\left(x,u\right)-\underline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}\right)\\
& &&\subseteq \left(\underline{y}^*,+\infty\right),
\end{aligned} \tag{A.18}$$

and that

$$\begin{aligned}
&\alpha\left(\overline{x},\overline{u}\right) &&\in \left(\underline{y}^*,\frac{m_1^Y\left(x,u\right)-\underline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}\right)\\
\Leftrightarrow\ &\gamma\left(\overline{x},\overline{u}\right) &&\in \left(\underline{y}^*,+\infty\right).
\end{aligned} \tag{A.19}$$

A.5 Proof of Proposition 12

This proof is essentially the same proof of proposition 11 under assumption 7.3.(a) (appendix A.4.1). Fix $\overline{u}\in[0,1]$, $\overline{x}\in\mathcal{X}$ and $\delta\left(\overline{x},\overline{u}\right)\in\mathbb{R}$ arbitrarily. For brevity, define $\alpha\left(\overline{x},\overline{u}\right):=\delta\left(\overline{x},\overline{u}\right)+\dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}$ and $\gamma\left(\overline{x},\overline{u}\right):=\dfrac{m_1^Y\left(\overline{x},\overline{u}\right)-\alpha\left(\overline{x},\overline{u}\right)\cdot m_0^S\left(\overline{x},\overline{u}\right)}{\Delta_S\left(\overline{x},\overline{u}\right)}$. Note that $\alpha\left(\overline{x},\overline{u}\right)\in\mathbb{R}=\mathcal{Y}^*$ and $\gamma\left(\overline{x},\overline{u}\right)\in\mathbb{R}=\mathcal{Y}^*$.

I define the random variables $\left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}\right)$ using the joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ described by steps 1-12 in Appendix A.4.1 for the case of convex support $\mathcal{Y}^*$. Note that equation (19) is trivially true when $\mathcal{Y}^*=\mathbb{R}$. Moreover, equations (18) and (20) are valid by the argument described in steps 13-17 in Appendix A.4.1.

I can, then, conclude that proposition 12 is true. ■

A.6 Proof of Equations (22) and (23)

I first prove that equation (22) holds. For any $A\in\{Y,S\}$, observe that

$$\begin{aligned}
\mathbb{E}\left[A\,|X=x,P\left(W\right)=p,D=0\right]&=\mathbb{E}\left[A_0\,|X=x,P\left(W\right)=p,D=0\right]\\
&=\mathbb{E}\left[A_0\,|X=x,P\left(W\right)=p,P\left(W\right)<U\right]
\end{aligned}$$

by equation (1)

$$= \mathbb{E}\left[A_0 \,|X = x, P\left(W\right) = p, p < U\right]$$

$$= \mathbb{E}\left[A_0 \,|X = x, p < U\right]$$

by assumption (1)

$$= \frac{\mathbb{E}\left[\mathbf{1}\left\{p < U\right\} \cdot A_0 \,|X = x\right]}{\mathbb{P}\left[p < U \,|X = x\right]}$$

by the definition of conditional expectation

$$= \frac{\mathbb{E}\left[\mathbf{1}\left\{p < U\right\} \cdot A_0 \,|X = x\right]}{1 - p}$$

by the normalization $U\,|X \sim \text{Uniform}\left[0, 1\right]$

$$= \frac{\mathbb{E}\left[\mathbf{1}\left\{p < U\right\} \cdot \mathbb{E}\left[A_0 \,|X = x, U = u\right] |X = x\right]}{1 - p}$$

by the Law of Iterated Expectations

$$= \frac{\int_p^1 m_0^A\left(x, u\right) \mathrm{d}u}{1 - p}$$

by the normalization $U\,|X \sim \text{Uniform}\left[0, 1\right]$,

implying that

$$\begin{aligned}
\frac{\partial \mathbb{E}\left[A\,|X = x, P\left(W\right) = p, D = 0\right]}{\partial p} &= \frac{-m_0^A\left(x, p\right)}{1 - p} + \frac{\mathbb{E}\left[\mathbf{1}\left\{p < U\right\} \cdot A_0 \,|X = x\right]}{\left(1 - p\right)^2} \\
&= \frac{-m_0^A\left(x, p\right)}{1 - p} + \frac{\mathbb{E}\left[\mathbf{1}\left\{p < U\right\} \cdot A_0 \,|X = x\right]}{\left(1 - p\right) \cdot \mathbb{P}\left[p < U \,|X = x\right]} \\
&\qquad \text{by the normalization } U\,|X \sim \text{Uniform}\left[0, 1\right] \\
&= \frac{-m_0^A\left(x, p\right)}{1 - p} + \frac{\mathbb{E}\left[A\,|X = x, P\left(W\right) = p, D = 0\right]}{1 - p}
\end{aligned}$$

Rearranging the last expression, I can derive equation (22):

$$\begin{aligned}
m_0^A\left(x, p\right) = {} & \mathbb{E}\left[A\,|X = x, P\left(W\right) = p, D = 0\right] \\
& - \frac{\partial \mathbb{E}\left[A\,|X = x, P\left(W\right) = p, D = 0\right]}{\partial p} \cdot \left(1 - p\right).
\end{aligned}$$

Equation (23) is derived in an analogous way using $\mathbb{E}\left[A \,|X=x, P\left(W\right)=p, D=1\right]$ and its derivative with respect to the propensity score. ■

A.7 Proof of Equations (33) and (34)

We first prove that equation (33) holds. For any $A \in \{Y, S\}$, observe that

$$\begin{aligned}\mathbb{E}\left[A \,|X=x, P\left(W\right)=p_n, D=0\right] &= \frac{\int_{p_n}^{1} m_0^A\left(x,u\right)\,\mathrm{d}u}{1-p_n} \\ &\quad \text{according to Appendix } A.6 \\ &= \frac{\int_{p_n}^{1} M^A\left(u, \boldsymbol{\theta}_{x,0}^A\right)\,\mathrm{d}u}{1-p_n} \\ &\quad \text{by equation (31).}\end{aligned}$$

Equation (34) is derived in an analogous way using $\mathbb{E}\left[A \,|X=x, P\left(W\right)=p_n, D=1\right]$. ■

B Numerical Example: MTE bounds (corollary 10) can be informative

Here, I provide a numerical example in which the bounds in equations (13) and (14) are informative (i.e., tighter than $\overline{y}^* - \underline{y}^*$) for any value of the latent variable U.

Let $\underline{y}^* = 0$ and $\overline{y}^* = 1$. Assume that there is no covariate and that, for any $u \in [0, 1]$,

$$\begin{aligned}
m_0^S(u) &= \frac{13}{16} - \frac{1}{16}u \\
m_1^S(u) &= \frac{15}{16} - \frac{2}{16}u \\
\mathbb{E}\left[Y_1^* \left| U = u, S_0 = 1, S_1 = 1\right.\right] &= \frac{5}{16} - \frac{4}{16}u \\
\Delta_Y^{NO}(u) &= \frac{9}{16} - \frac{2}{16}u,
\end{aligned}$$

implying that

$$\begin{aligned}
\Delta_S(u) &= \frac{2}{16} - \frac{1}{16}u \\
m_1^Y(u) &= \frac{83 - 70u + 6u^2}{256}.
\end{aligned}$$

Moreover, assume that $m_0^Y(u) = \frac{10}{16} - \frac{1}{16}u$, implying that

$$\mathbb{E}\left[Y_1^* \left| U = u, S_0 = 1, S_1 = 1\right.\right] = \frac{-10 + u}{-13 + u}.$$

Consequently, the target parameter is

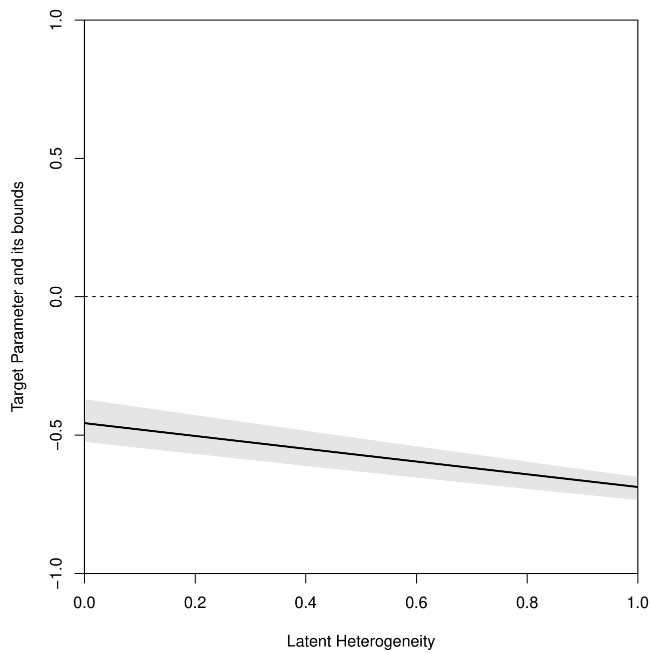
$$\Delta_{Y^*}^{OO}(u) = \frac{95 + 41u - 4u^2}{-208 + 16u}$$

and the lower and upper bounds in equations (13) and (14) are, respectively, given by

$$\begin{aligned}
\underline{\Delta_{Y^*}^{OO}}(u) &= \frac{109 + 38u - 6u^2}{-208 + 16u}, \text{ and} \\
\overline{\Delta_{Y^*}^{OO}}(u) &= \frac{77 + 54u - 6u^2}{-208 + 16u}.
\end{aligned}$$

Figure B.1 plots the target parameter, $\Delta_{Y^*}^{OO}(u)$, as a solid black line and the lower and upper bounds, $\underline{\Delta_{Y^*}^{OO}}(u)$ and $\overline{\Delta_{Y^*}^{OO}}(u)$, as a gray area. Observe that the proposed bounds are close to the target parameter in the entire unit interval. As a consequence, this numerical example illustrates that the bounds in equations (13) and (14) can be informative about the target parameter.

Figure B.1: Target Parameter and Its Bounds



C MTE bounds under a Mean Dominance Assumption

In the literature about partial identification of treatment effects (Huber & Mellace (2015), Huber et al. (2017)), mean or stochastic dominance assumptions are used to sharpen the bounds on the parameter of interest. Here, I use one type of mean dominance assumption to sharpen the bounds on $\Delta_{Y^*}^{OO}$ given by corollary 10. In particular, I assume either

Assumption C.1.A *The potential outcome when treated for the always-observed sub-population is greater than or equal to the same parameter for the observed-only-when-treated sub-population:*

$$\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \geq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 0, S_1 = 1\right.\right]$$

for any $x \in \mathcal{X}$ and $u \in [0, 1]$

or

Assumption C.1.B *The potential outcome when treated for the always-observed sub-population is less than or equal to the same parameter for the observed-only-when-treated sub-population:*

$$\mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 1, S_1 = 1\right.\right] \leq \mathbb{E}\left[Y_1^* \left| X = x, U = u, S_0 = 0, S_1 = 1\right.\right]$$

for any $x \in \mathcal{X}$ and $u \in [0, 1]$.

Note that assumption C.1.A implies that $\Delta_{Y}^{NO}(x, u) \leq m_1^Y(x, u)$ by equations (A.4) and (A.5), while assumption C.1.B implies that $\Delta_{Y}^{NO}(x, u) \geq m_1^Y(x, u)$. As a consequence, by following the same steps of the proof of corollary 10, I can derive:

Corollary C.2.A *Fix $u \in [0, 1]$ and $x \in \mathcal{X}$ arbitrarily. Suppose that the $m_0^Y(x, u)$, $m_1^Y(x, u)$, $m_0^S(x, u)$ and $\Delta_S(x, u)$ are point-identified. Define $\kappa := \min\left\{m_1^Y(x, u), \overline{y}^*\right\}$.*

Under assumptions 1-6, 7.1, 8 and C.1.A, the bounds on $\Delta_{Y^*}^{OO}(x, u)$ are given by

$$\Delta_{Y^*}^{OO}(x, u) \geq \max\left\{\frac{m_1^Y(x, u) - \kappa \cdot \Delta_S(x, u)}{m_0^S(x, u)}, \underline{y}^*\right\} - \frac{m_0^Y(x, u)}{m_0^S(x, u)} =: \underline{\Delta_{Y^*}^{OO}}(x, u) \qquad \text{(C.1)}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{C.2}$$

Under assumptions 1-6, 7.2, 8 and C.1.A, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ are given by

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{m_1^Y(x,u) - \kappa \cdot \Delta_S(x,u)}{m_0^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{C.3}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \overline{y}^* - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{C.4}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)), 8 and C.1.A, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ are given by

$$\Delta_{Y^*}^{OO}(x,u) \geq \max\left\{\frac{m_1^Y(x,u) - \kappa \cdot \Delta_S(x,u)}{m_0^S(x,u)}, \underline{y}^*\right\} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{C.5}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \min\left\{\frac{m_1^Y(x,u) - \underline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)}, \overline{y}^*\right\} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{C.6}$$

and

Corollary C.2.B *Fix $u \in [0,1]$ and $x \in \mathcal{X}$ arbitrarily. Suppose that the $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$ and $\Delta_S(x,u)$ are point-identified. Define $\kappa := \max\left\{m_1^Y(x,u), \underline{y}^*\right\}$.*

Under assumptions 1-6, 7.1, 8 and C.1.B, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ are given by

$$\Delta_{Y^*}^{OO}(x,u) \geq \underline{y}^* - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{C.7}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \frac{m_1^Y(x,u) - \kappa \cdot \Delta_S(x,u)}{m_0^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{C.8}$$

Under assumptions 1-6, 7.2, 8 and C.1.B, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ are given by

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{m_1^Y(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{C.9}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \min\left\{\frac{m_1^Y(x,u) - \kappa \cdot \Delta_S(x,u)}{m_0^S(x,u)}, \overline{y}^*\right\} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{C.10}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)), 8 and C.1.B, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ are given by

$$\Delta_{Y^*}^{OO}(x,u) \geq \max\left\{\frac{m_1^Y(x,u) - \overline{y}^* \cdot \Delta_S(x,u)}{m_0^S(x,u)}, \underline{y}^*\right\} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \underline{\Delta_{Y^*}^{OO}}(x,u) \tag{C.11}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \min\left\{\frac{m_1^Y(x,u) - \kappa \cdot \Delta_S(x,u)}{m_0^S(x,u)}, \overline{y}^*\right\} - \frac{m_0^Y(x,u)}{m_0^S(x,u)} =: \overline{\Delta_{Y^*}^{OO}}(x,u). \tag{C.12}$$

The bounds in corollaries C.2.A and C.2.B can be identified using the strategies that were described in sections 4 and 5.

D Bounds on the MTE for the Observed-only-when-treated Sub-population

Here, I use the same notation of section 3 and I am interested in the following target parameter: $m_1^{NO}(x,u) \coloneqq \mathbb{E}\left[Y_1^* \left| X=x, U=u, S_0=0, S_1=1\right.\right]$. Following the same steps of the proof of proposition 9, I can show that:

Proposition D.1 *Suppose that the $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$ and $\Delta_S(x,u)$ are point-identified.*

Under assumptions 1-6, 7.1 and 8, the bounds on $m_1^{NO}(x,u)$ are given by

$$\underline{m_1^{NO}}(x,u) \coloneqq \underline{y}^* \leq m_1^{NO}(x,u) \leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot m_0^S(x,u)}{\Delta_S(x,u)} \eqqcolon \overline{m_1^{NO}}(x,u). \tag{D.1}$$

Under assumptions 1-6, 7.2 and 8, the bounds on $m_1^{NO}(x,u)$ are given by

$$\underline{m_1^{NO}}(x,u) \coloneqq \frac{m_1^Y(x,u) - \overline{y}^* \cdot m_0^S(x,u)}{\Delta_S(x,u)} \leq m_1^{NO}(x,u) \leq \overline{y}^* \eqqcolon \overline{m_1^{NO}}(x,u). \tag{D.2}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)) and 8, the bounds on $m_1^{NO}(x,u)$ are given by

$$\underline{m_1^{NO}}(x,u) \coloneqq \frac{m_1^Y(x,u) - \overline{y}^* \cdot m_0^S(x,u)}{\Delta_S(x,u)} \leq m_1^{NO}(x,u) \leq \frac{m_1^Y(x,u) - \underline{y}^* \cdot m_0^S(x,u)}{\Delta_S(x,u)} \eqqcolon \overline{m_1^{NO}}(x,u). \tag{D.3}$$

Following the same proof of proposition 11 (see Remark 2 at the end of Appendix A.4.1), I can also show that:

Proposition D.2 *Suppose that the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified at every pair $(x,u) \in \mathcal{X} \times [0,1]$. Under assumptions 1-6, 7 (sub-cases 1, 2, 3(a) or 3(b)) and 8, the bounds $\underline{m_1^{NO}}$ and $\overline{m_1^{NO}}$, given by proposition D.1, are point-wise sharp, i.e., for any $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\gamma(\overline{x},\overline{u}) \in \left(\underline{m_1^{NO}}(\overline{x},\overline{u}), \overline{m_1^{NO}}(\overline{x},\overline{u})\right)$, there exist random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ such that*

$$\tilde{m}_1^{NO}(\overline{x},\overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* \middle| X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=0, \tilde{S}_1=1\right] = \gamma(\overline{x},\overline{u}), \tag{D.4}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{D.5}$$

and

$$F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z, X}(y, d, s, z, \overline{x}) = F_{Y,D,S,Z,X}(y, d, s, z, \overline{x}) \tag{D.6}$$

for any $(y, d, s, z) \in \mathbb{R}^4$, *where* $\tilde{D} := \mathbf{1}\left\{P(X, Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0, X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1, X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

Finally, following the same proof of proposition 12, I can also show that:

Proposition D.3 *Suppose that the functions* m_0^Y, m_1^Y, m_0^S *and* Δ_S *are point-identified at every pair* $(x, u) \in \mathcal{X} \times [0,1]$. *Impose assumptions 1-6 and 8. If* $\mathcal{Y}^* = \mathbb{R}$, *then, for any* $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ *and* $\gamma(\overline{x}, \overline{u}) \in \mathbb{R}$, *there exist random variables* $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ *such that*

$$\tilde{m}_1^{NO}(\overline{x}, \overline{u}) := \mathbb{E}\left[\tilde{Y}_1^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 0, \tilde{S}_1 = 1\right] = \gamma(\overline{x}, \overline{u}), \tag{D.7}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{D.8}$$

and

$$F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z, X}(y, d, s, z, \overline{x}) = F_{Y,D,S,Z,X}(y, d, s, z, \overline{x}) \tag{D.9}$$

for any $(y, d, s, z) \in \mathbb{R}^4$, *where* $\tilde{D} := \mathbf{1}\left\{P(X, Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0, X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1, X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

E Negative Treatment Effect on the Selection Indicator

Even when sample selection is monotone (equation (2)), assumption 8 may be invalid in some empirical applications. In particular, it might be the case that the following assumption holds:

Assumption E.1 *Treatment has a negative effect on the sample selection indicator for all individuals, i.e.,* $Q(0,x) > Q(1,x) > 0$ *for any* $x \in \mathcal{X}$.

I stress that this assumption is testable according to Machado et al. (2018).

With obvious modifications to the proofs of corollary 10 and propositions 11 and 12 (see the proofs of propositions F.3 and F.4), I can show that the target parameter in section 3 can be bounded, that its bounds are sharp and that it is impossible to derive bounds for the target parameter with only assumptions 1-6 and E.1. First, I state a result that is analogous to corollary 10.

Proposition E.2 *Fix* $u \in [0,1]$ *and* $x \in \mathcal{X}$ *arbitrarily. Suppose that the* $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$ *and* $\Delta_S(x,u)$ *are point-identified.*

Under assumptions 1-6, 7.1 and E.1, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ *are given by*

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \frac{m_0^Y(x,u) - \underline{y}^* \cdot (-\Delta_S(x,u))}{m_1^S(x,u)} =: \underline{\Lambda_{Y^*}^{OO}}(x,u) \tag{E.1}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \underline{y}^* =: \overline{\Lambda_{Y^*}^{OO}}(x,u). \tag{E.2}$$

Under assumptions 1-6, 7.2 and E.1, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ *are given by*

$$\Delta_{Y^*}^{OO}(x,u) \geq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \overline{y}^* =: \underline{\Lambda_{Y^*}^{OO}}(x,u) \tag{E.3}$$

and that

$$\Delta_{Y^*}^{OO}(x,u) \leq \frac{m_1^Y(x,u)}{m_1^S(x,u)} - \frac{m_0^Y(x,u) - \overline{y}^* \cdot (-\Delta_S(x,u))}{m_1^S(x,u)} =: \overline{\Lambda_{Y^*}^{OO}}(x,u). \tag{E.4}$$

Under assumptions 1-6, 7.3 (sub-case (a) or (b)) and E.1, the bounds on $\Delta^{OO}_{Y^*}(x,u)$ *are given by*

$$\Delta^{OO}_{Y^*}(x,u) \geq \frac{m^Y_1(x,u)}{m^S_1(x,u)} - \min\left\{\frac{m^Y_0(x,u) - \underline{y}^* \cdot (-\Delta_S(x,u))}{m^S_1(x,u)}, \overline{y}^*\right\} =: \underline{\Lambda^{OO}_{Y^*}}(x,u) \tag{E.5}$$

and that

$$\Delta^{OO}_{Y^*}(x,u) \leq \frac{m^Y_1(x,u)}{m^S_1(x,u)} - \max\left\{\frac{m^Y_0(x,u) - \overline{y}^* \cdot (-\Delta_S(x,u))}{m^S_1(x,u)}, \underline{y}^*\right\} =: \overline{\Lambda^{OO}_{Y^*}}(x,u)\,. \tag{E.6}$$

Second, I state a result that is analogous to proposition 11.

Proposition E.3 *Suppose that the functions* m^Y_0*,* m^Y_1*,* m^S_0 *and* Δ_S *are point-identified at every pair* $(x,u) \in \mathcal{X} \times [0,1]$*. Under assumptions 1-6, 7 (sub-cases 1, 2, 3(a) or 3(b)) and E.1, the bounds* $\underline{\Lambda^{OO}_{Y^*}}$ *and* $\overline{\Lambda^{OO}_{Y^*}}$*, given by proposition E.2, are point-wise sharp, i.e., for any* $\overline{u} \in [0,1]$*,* $\overline{x} \in \mathcal{X}$ *and* $\delta(\overline{x},\overline{u}) \in \left(\underline{\Lambda^{OO}_{Y^*}}(\overline{x},\overline{u}), \overline{\Lambda^{OO}_{Y^*}}(\overline{x},\overline{u})\right)$*, there exist random variables* $\left(\tilde{Y}^*_0, \tilde{Y}^*_1, \tilde{U}, \tilde{V}\right)$ *such that*

$$\Delta^{OO}_{\tilde{Y}^*}(\overline{x},\overline{u}) := \mathbb{E}\left[\tilde{Y}^*_1 - \tilde{Y}^*_0 \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x},\overline{u})\,, \tag{E.7}$$

$$\mathbb{P}\left[\left(\tilde{Y}^*_0, \tilde{Y}^*_1, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1]\,, \tag{E.8}$$

and

$$F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) = F_{Y,D,S,Z,X}(y,d,s,z,\overline{x}) \tag{E.9}$$

for any $(y,d,s,z) \in \mathbb{R}^4$*, where* $\tilde{D} := \mathbf{1}\left\{P(X,Z) \geq \tilde{U}\right\}$*,* $\tilde{S}_0 = \mathbf{1}\left\{Q(0,X) \geq \tilde{V}\right\}$*,* $\tilde{S}_1 = \mathbf{1}\left\{Q(1,X) \geq \tilde{V}\right\}$*,* $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}^*_0$*,* $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}^*_1$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$*.*

Finally, I state a result that is analogous to proposition 12.

Proposition E.4 *Suppose that the functions* m^Y_0*,* m^Y_1*,* m^S_0 *and* Δ_S *are point-identified at every pair* $(x,u) \in \mathcal{X} \times [0,1]$*. Impose assumptions 1-6 and E.1. If* $\mathcal{Y}^* = \mathbb{R}$*, then, for any*

$\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x}, \overline{u}) \in \mathbb{R}$, there exist random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ such that

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x}, \overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x}, \overline{u}), \tag{E.10}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{E.11}$$

and

$$F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z, X}(y, d, s, z, \overline{x}) = F_{Y, D, S, Z, X}(y, d, s, z, \overline{x}) \tag{E.12}$$

for any $(y, d, s, z) \in \mathbb{R}^4$, where $\tilde{D} \coloneqq \mathbf{1}\left\{P(X, Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0, X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1, X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ and $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

F Monotone Sample Selection

Depending on the results of the test proposed by Machado et al. (2018), a researcher may want to be agnostic about the direction of the monotone selection problem and impose only equation (2), while ruling out uninteresting cases. In such situation, it is reasonable to assume:

Assumption F.1 *Treatment has a monotone effect on the sample selection indicator for all individuals, i.e., **either** (i) $Q(1,x) > Q(0,x) > 0$ for any $x \in \mathcal{X}$ **or** (ii) $Q(0,x) > Q(1,x) > 0$ for any $x \in \mathcal{X}$.*

I stress that assumption F.1 only strengthens equation (2) by ruling out the theoretically uninteresting cases mention below assumption (8).

By combining corollary 10 and proposition E.2, I find that:

Proposition F.2 *Fix $u \in [0,1]$ and $x \in \mathcal{X}$ arbitrarily. Suppose that the $m_0^Y(x,u)$, $m_1^Y(x,u)$, $m_0^S(x,u)$ and $\Delta_S(x,u)$ are point-identified. Under assumptions 1-6, 7 and F.1, the bounds on $\Delta_{Y^*}^{OO}(x,u)$ are given by*

$$\begin{aligned}\underline{\Upsilon_{Y^*}^{OO}}(x,u) &\coloneqq \min\left\{\underline{\Delta_{Y^*}^{OO}}(x,u), \underline{\Lambda_{Y^*}^{OO}}(x,u)\right\} \\ &\leq \Delta_{Y^*}^{OO}(x,u) \\ &\leq \max\left\{\overline{\Delta_{Y^*}^{OO}}(x,u), \overline{\Lambda_{Y^*}^{OO}}(x,u)\right\} \eqqcolon \overline{\Upsilon_{Y^*}^{OO}}(x,u)\end{aligned} \tag{F.1}$$

Most importantly, those bounds are also point-wise sharp:[11]

Proposition F.3 *Suppose that the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified at every pair $(x,u) \in \mathcal{X} \times [0,1]$. Under assumptions 1-6, 7 (sub-cases 1, 2, 3(a) or 3(b)) and F.1, the bounds $\underline{\Upsilon_{Y^*}^{OO}}$ and $\overline{\Upsilon_{Y^*}^{OO}}$, given by corollary 10, are point-wise sharp, i.e., for any $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \left(\underline{\Upsilon_{Y^*}^{OO}}(\overline{x},\overline{u}), \overline{\Upsilon_{Y^*}^{OO}}(\overline{x},\overline{u})\right)$, there exist random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$*

[11] The proof of propositions F.3 and F.4 are located at the end of Appendix F.

such that

$$\Delta^{OO}_{\tilde{Y}^*}(\overline{x},\overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}^*_1 - \tilde{Y}^*_0 \middle| X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] = \delta(\overline{x},\overline{u}), \tag{F.2}$$

$$\mathbb{P}\left[\left(\tilde{Y}^*_0, \tilde{Y}^*_1, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X=\overline{x}, \tilde{U}=u\right] = 1 \text{ for any } u \in [0,1], \tag{F.3}$$

and

$$F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) = F_{Y,D,S,Z,X}(y,d,s,z,\overline{x}) \tag{F.4}$$

for any $(y,d,s,z) \in \mathbb{R}^4$, *where* $\tilde{D} \coloneqq \mathbf{1}\left\{P(X,Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0,X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1,X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}^*_0$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}^*_1$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1-\tilde{D}\right) \cdot \tilde{Y}_0$.

Finally, I state an impossibility result that is analogous to proposition 12.

Proposition F.4 *Suppose that the functions* m^Y_0, m^Y_1, m^S_0 *and* Δ_S *are point-identified at every pair* $(x,u) \in \mathcal{X} \times [0,1]$. *Impose assumptions 1-6 and F.1. If* $\mathcal{Y}^* = \mathbb{R}$, *then, for any* $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ *and* $\delta(\overline{x},\overline{u}) \in \mathbb{R}$, *there exist random variables* $\left(\tilde{Y}^*_0, \tilde{Y}^*_1, \tilde{U}, \tilde{V}\right)$ *such that*

$$\Delta^{OO}_{\tilde{Y}^*}(\overline{x},\overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}^*_1 - \tilde{Y}^*_0 \middle| X=\overline{x}, \tilde{U}=\overline{u}, \tilde{S}_0=1, \tilde{S}_1=1\right] = \delta(\overline{x},\overline{u}), \tag{F.5}$$

$$\mathbb{P}\left[\left(\tilde{Y}^*_0, \tilde{Y}^*_1, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X=\overline{x}, \tilde{U}=u\right] = 1 \text{ for any } u \in [0,1], \tag{F.6}$$

and

$$F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) = F_{Y,D,S,Z,X}(y,d,s,z,\overline{x}) \tag{F.7}$$

for any $(y,d,s,z) \in \mathbb{R}^4$, *where* $\tilde{D} \coloneqq \mathbf{1}\left\{P(X,Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0,X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1,X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}^*_0$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}^*_1$ *and* $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1-\tilde{D}\right) \cdot \tilde{Y}_0$.

Proof of Proposition F.3. I only prove proposition F.3 under assumption 7.3 (subcases (a) and (b)).The proofs of proposition F.3 under assumptions 7.1 and 7.2 are trivial modifications of the proof presented below.

Fix $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \left(\underline{\Upsilon^{OO}_{Y^*}}(\overline{x},\overline{u}), \overline{\Upsilon^{OO}_{Y^*}}(\overline{x},\overline{u})\right)$ arbitrarily. For brevity, define

$$\alpha\left(\overline{x},\overline{u}\right):=\mathbf{1}\left\{Q\left(1,x\right)>Q\left(0,x\right)\right\}\cdot\left(\delta\left(\overline{x},\overline{u}\right)+\frac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}\right)$$
$$+\mathbf{1}\left\{Q\left(1,x\right)<Q\left(0,x\right)\right\}\cdot\left(-\delta\left(\overline{x},\overline{u}\right)+\frac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}\right),$$

$$\gamma\left(\overline{x},\overline{u}\right):=\mathbf{1}\left\{Q\left(1,x\right)>Q\left(0,x\right)\right\}\cdot\left(\frac{m_1^Y\left(\overline{x},\overline{u}\right)-\alpha\left(\overline{x},\overline{u}\right)\cdot m_0^S\left(\overline{x},\overline{u}\right)}{\Delta_S\left(\overline{x},\overline{u}\right)}\right)$$
$$+\mathbf{1}\left\{Q\left(1,x\right)<Q\left(0,x\right)\right\}\cdot\left(\frac{m_0^Y\left(\overline{x},\overline{u}\right)-\alpha\left(\overline{x},\overline{u}\right)\cdot m_1^S\left(\overline{x},\overline{u}\right)}{-\Delta_S\left(\overline{x},\overline{u}\right)}\right),$$

$$\underline{Q}\left(x\right)=\min\left\{Q\left(0,x\right),Q\left(1,x\right)\right\},$$

and

$$\overline{Q}\left(x\right)=\max\left\{Q\left(0,x\right),Q\left(1,x\right)\right\}.$$

Note that

$$\alpha\left(\overline{x},\overline{u}\right)\in\left(\underline{y}^*,\overline{y}^*\right),\tag{F.8}$$

and that

$$\gamma\left(\overline{x},\overline{u}\right)\in\left(\underline{y}^*,\overline{y}^*\right).\tag{F.9}$$

The strategy of this proof consists of defining random variables $\left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}\right)$ through their joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ and, then, checking that equations (F.2), (F.3) and (F.4) are satisfied. I fix $(y_0,y_1,u,v,z,x)\in\mathbb{R}^6$ and define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ in twelve steps:

Step 1. For $x\notin\mathcal{X}$, $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0,y_1,u,v,z,x\right)=F_{Y_0^*,Y_1^*,U,V,Z,X}\left(y_0,y_1,u,v,z,x\right)$.

Step 2. From now on, assume that $x\in\mathcal{X}$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0,y_1,u,v,z,x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\left|x\right.\right)\cdot F_X\left(x\right),$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\,|x\right)$. Moreover, I impose

$$Z \perp\!\!\!\perp \left(\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}\right)|X$$

by writing

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\,|x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)\cdot F_{Z|X}\left(z\,|x\right),$$

implying that it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 3. For $u\notin[0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)=F_{Y_0^*,Y_1^*,U,V|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 4. From now on, assume that $u\in[0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)\cdot F_{\tilde{U}|X}\left(u\,|x\right),$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)$ and $F_{\tilde{U}|X}\left(u\,|x\right)$.

Step 5. I define $F_{\tilde{U}|X}\left(u\,|x\right)=F_{U|X}\left(u\,|x\right)=u$.

Step 6. For any $u\neq\overline{u}$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)=F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,u\right)$.

Step 7. For any $v\notin[0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,\overline{u}\right)=F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,\overline{u}\right)$.

Step 8. From now on, assume that $v\in[0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,\overline{u}\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)\cdot F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right),$$

it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)$ and $F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)$.

Step 9. I define

$$F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)=\begin{cases} m_0^S\left(x,\overline{u}\right)\cdot\dfrac{v}{\underline{Q}\left(x\right)} & \text{if } v\leq\underline{Q}\left(x\right) \\ m_0^S\left(x,\overline{u}\right)+\Delta_S\left(x,\overline{u}\right)\cdot\dfrac{v-\underline{Q}\left(x\right)}{\overline{Q}\left(x\right)-\underline{Q}\left(x\right)} & \text{if } \underline{Q}\left(x\right)<v\leq\overline{Q}\left(x\right) \\ m_1^S\left(x,\overline{u}\right)+\left(1-m_1^S\left(x,\overline{u}\right)\right)\dfrac{v-\overline{Q}\left(x\right)}{1-\overline{Q}\left(x\right)} & \text{if } \overline{Q}\left(x\right)<v \end{cases}.$$

Step 10. I write $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)=F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right)\cdot F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\,|x,\overline{u},v\right)$, implying that I can separately define $F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right)$ and $F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\,|x,\overline{u},v\right)$.

Step 11. When $Q\left(1,x\right)>Q\left(0,x\right)$ and $\mathcal{Y}^*$ is a bounded interval (sub-case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right)=\begin{cases} \mathbf{1}\left\{y_0\geq\dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}\right\} & \text{if } v\leq\underline{Q}\left(x\right) \\ ------------ & -------- \\ \mathbf{1}\left\{y_0\geq\dfrac{\underline{y}^*+\overline{y}^*}{2}\right\} & \text{if } \underline{Q}\left(x\right)<v \end{cases}.$$

When $Q\left(1,x\right)>Q\left(0,x\right)$ and $\overline{y}^*=\max\left\{y\in\mathcal{Y}^*\right\}$ and $\underline{y}^*=\min\left\{y\in\mathcal{Y}^*\right\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right)=\begin{cases} 0 & \text{if } y_0<\underline{y}^* \text{ and } v\leq\underline{Q}\left(x\right) \\ 1-\dfrac{\dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}-\underline{y}^*}{\overline{y}^*-\underline{y}^*} & \text{if } \underline{y}^*\leq y_0<\overline{y}^* \text{ and } v\leq\underline{Q}\left(x\right) \\ 1 & \text{if } \overline{y}^*\leq y_0 \text{ and } v\leq\underline{Q}\left(x\right) \\ ------------ & ---------------- \\ \mathbf{1}\left\{y_0\geq\overline{y}^*\right\} & \text{if } \underline{Q}\left(x\right)<v \end{cases}.$$

which are valid cumulative distribution functions because $\dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} \in \left[\underline{y}^*, \overline{y}^*\right]$.

When $Q(1,x) < Q(0,x)$ and $\mathcal{Y}^*$ is a bounded interval (case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}(y_0\,|x,\overline{u},v) = \begin{cases} \mathbf{1}\{y_0 \geq \alpha(\overline{x},\overline{u})\} & \text{if } v \leq \underline{Q}(x) \\ -------- & ------------ \\ \mathbf{1}\{y_0 \geq \gamma(\overline{x},\overline{u})\} & \text{if } \underline{Q}(x) < v \leq \overline{Q}(x) \\ -------- & ------------ \\ \mathbf{1}\left\{y_0 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2}\right\} & \text{if } \overline{Q}(x) < v \end{cases}.$$

When $Q(1,x) < Q(0,x)$ and $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}(y_0\,|x,\overline{u},v) = \begin{cases} 0 & \text{if } y_0 < \underline{y}^* \text{ and } v \leq \underline{Q}(x) \\ 1 - \dfrac{\alpha(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_0 < \overline{y}^* \text{ and } v \leq \underline{Q}(x) \\ 1 & \text{if } \overline{y}^* \leq y_0 \text{ and } v \leq \underline{Q}(x) \\ -------- & ------------------ \\ 0 & \text{if } y_0 < \underline{y}^* \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ 1 - \dfrac{\gamma(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_0 < \overline{y}^* \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ 1 & \text{if } \overline{y}^* \leq y_0 \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ -------- & ------------------ \\ \mathbf{1}\{y_0 \geq \overline{y}^*\} & \text{if } \overline{Q}(x) < v \end{cases}.$$

which are valid cumulative distribution functions because of equations (F.8) and (F.9).

Step 12. When $Q(1,x) > Q(0,x)$ and $\mathcal{Y}^*$ is a bounded interval (case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_1|x,\overline{u},v) = \begin{cases} \mathbf{1}\{y_1 \geq \alpha(\overline{x},\overline{u})\} & \text{if } v \leq \underline{Q}(x) \\ -------- & ----------- \\ \mathbf{1}\{y_1 \geq \gamma(\overline{x},\overline{u})\} & \text{if } \underline{Q}(x) < v \leq \overline{Q}(x) \\ -------- & ----------- \\ \mathbf{1}\left\{y_1 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2}\right\} & \text{if } \overline{Q}(x) < v \end{cases}.$$

When $Q(1,x) > Q(0,x)$ and $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_1|x,\overline{u},v) = \begin{cases} 0 & \text{if } y_1 < \underline{y}^* \text{ and } v \leq \underline{Q}(x) \\ 1 - \dfrac{\alpha(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } v \leq \underline{Q}(x) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } v \leq \underline{Q}(x) \\ -------- & ------------------ \\ 0 & \text{if } y_1 < \underline{y}^* \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ 1 - \dfrac{\gamma(\overline{x},\overline{u}) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } \underline{Q}(x) < v \leq \overline{Q}(x) \\ -------- & ------------------ \\ \mathbf{1}\{y_1 \geq \overline{y}^*\} & \text{if } \overline{Q}(x) < v \end{cases}.$$

which are valid cumulative distribution functions because of equations (A.10) and (A.11).

When $Q(1,x) < Q(0,x)$ and $\mathcal{Y}^*$ is a bounded interval (sub-case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\left|x,\overline{u},v\right.\right)=\begin{cases}\mathbf{1}\left\{y_1\geq\dfrac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}\right\} & \text{if } v\leq\underline{Q}\left(x\right)\\ -------- & --------\\ \mathbf{1}\left\{y_1\geq\dfrac{\underline{y}^*+\overline{y}^*}{2}\right\} & \text{if } \underline{Q}\left(x\right)<v\end{cases}.$$

When $Q\left(1,x\right)<Q\left(0,x\right)$ and $\overline{y}^*=\max\left\{y\in\mathcal{Y}^*\right\}$ and $\underline{y}^*=\min\left\{y\in\mathcal{Y}^*\right\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\left|x,\overline{u},v\right.\right)=\begin{cases}0 & \text{if } y_1<\underline{y}^* \text{ and } v\leq\underline{Q}\left(x\right)\\ 1-\dfrac{\dfrac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}-\underline{y}^*}{\overline{y}^*-\underline{y}^*} & \text{if } \underline{y}^*\leq y_1<\overline{y}^* \text{ and } v\leq\underline{Q}\left(x\right)\\ 1 & \text{if } \overline{y}^*\leq y_1 \text{ and } v\leq\underline{Q}\left(x\right)\\ -------- & ------------\\ \mathbf{1}\left\{y_1\geq\overline{y}^*\right\} & \text{if } \underline{Q}\left(x\right)<v\end{cases}.$$

which are valid cumulative distribution functions because $\dfrac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}\in\left[\underline{y}^*,\overline{y}^*\right]$.

Having defined the joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$, note that equations (F.8) and (F.9), the facts $\dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}\in\left[\underline{y}^*,\overline{y}^*\right]$ and $\dfrac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}\in\left[\underline{y}^*,\overline{y}^*\right]$, and steps 7-12 ensure that equation (F.3) holds.

Now, I show, in three steps, that equation (F.2) holds.

Step 13. Observe that

$$\begin{aligned}&\mathbb{E}\left[\tilde{Y}_1^*\left|X=\overline{x},\tilde{U}=\overline{u},\tilde{S}_0=1,\tilde{S}_1=1\right.\right]\\ &\quad=\mathbf{1}\left\{Q\left(1,x\right)>Q\left(0,x\right)\right\}\cdot\alpha\left(\overline{x},\overline{u}\right)+\mathbf{1}\left\{Q\left(1,x\right)<Q\left(0,x\right)\right\}\cdot\frac{m_1^Y\left(\overline{x},\overline{u}\right)}{m_1^S\left(\overline{x},\overline{u}\right)}.\end{aligned}\tag{F.10}$$

Step 14. Notice that

$$\begin{aligned}
&\mathbb{E}\left[\tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] \\
&\quad = \mathbf{1}\left\{Q\left(1, x\right) > Q\left(0, x\right)\right\} \cdot \frac{m_0^Y\left(\overline{x}, \overline{u}\right)}{m_0^S\left(\overline{x}, \overline{u}\right)} + \mathbf{1}\left\{Q\left(1, x\right) < Q\left(0, x\right)\right\} \cdot \alpha\left(\overline{x}, \overline{u}\right). \quad \text{(F.11)}
\end{aligned}$$

Step 15. Note that Steps 13 and 14 imply that

$$\Delta_{\tilde{Y}^*}^{OO}\left(\overline{x}, \overline{u}\right) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta\left(\overline{x}, \overline{u}\right),$$

ensuring that equation (F.2) holds.

Finally, to show that equation (F.4) holds, it suffices to follow steps 16 and 17 in Appendix A.4.1.

I can, then, conclude that proposition F.3 is true. ■

Proof of Proposition F.4. This proof is essentially the same proof of proposition F.3 under assumption 7.3.(a). Fix $\overline{u} \in [0, 1]$, $\overline{x} \in \mathcal{X}$ and $\delta\left(\overline{x}, \overline{u}\right) \in \mathbb{R}$ arbitrarily. For brevity, define

$$\begin{aligned}
\alpha\left(\overline{x}, \overline{u}\right) \coloneqq\; &\mathbf{1}\left\{Q\left(1, x\right) > Q\left(0, x\right)\right\} \cdot \left(\delta\left(\overline{x}, \overline{u}\right) + \frac{m_0^Y\left(\overline{x}, \overline{u}\right)}{m_0^S\left(\overline{x}, \overline{u}\right)}\right) \\
&+ \mathbf{1}\left\{Q\left(1, x\right) < Q\left(0, x\right)\right\} \cdot \left(-\delta\left(\overline{x}, \overline{u}\right) + \frac{m_1^Y\left(\overline{x}, \overline{u}\right)}{m_1^S\left(\overline{x}, \overline{u}\right)}\right),
\end{aligned}$$

and

$$\begin{aligned}
\gamma\left(\overline{x}, \overline{u}\right) \coloneqq\; &\mathbf{1}\left\{Q\left(1, x\right) > Q\left(0, x\right)\right\} \cdot \left(\frac{m_1^Y\left(\overline{x}, \overline{u}\right) - \alpha\left(\overline{x}, \overline{u}\right) \cdot m_0^S\left(\overline{x}, \overline{u}\right)}{\Delta_S\left(\overline{x}, \overline{u}\right)}\right) \\
&+ \mathbf{1}\left\{Q\left(1, x\right) < Q\left(0, x\right)\right\} \cdot \left(\frac{m_0^Y\left(\overline{x}, \overline{u}\right) - \alpha\left(\overline{x}, \overline{u}\right) \cdot m_1^S\left(\overline{x}, \overline{u}\right)}{-\Delta_S\left(\overline{x}, \overline{u}\right)}\right).
\end{aligned}$$

Note that $\alpha\left(\overline{x}, \overline{u}\right) \in \mathbb{R} = \mathcal{Y}^*$ and $\gamma\left(\overline{x}, \overline{u}\right) \in \mathbb{R} = \mathcal{Y}^*$.

I define the random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ using the joint cumulative distribution function $F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z, X}$ described by steps 1-12 in the last proof for the case of convex support

$\mathcal{Y}^*$. Note that equation (F.6) is trivially true when $\mathcal{Y}^* = \mathbb{R}$. Moreover, equations (F.5) and (F.7) are valid by the argument described in the last proof

I can, then, conclude that proposition F.4 is true. $\blacksquare$

G Sharpness and Impossibility Results with Smoothness Restrictions

In the main text, I imposed no smoothness condition on the joint distribution of $(Y_0^*, Y_1^*, U, V, Z, X)$. Here, I impose the following smoothness condition:

Assumption G.1 *The conditional cumulative distribution functions $F_{V|X,U}$ are $F_{Y_0^*,Y_1^*|X,U,V}$ are continuous functions of U.*

As a consequence of this new assumption, propositions 11 and 12 have to be modified to accommodate infinitesimal violations of the data restriction and to ensure that the extra model restrictions imposed by assumption G.1 are also satisfied.

Proposition G.2 *Suppose that the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified at every pair $(x, u) \in \mathcal{X} \times [0,1]$. Under assumptions 1-6, 7 (sub-cases 1, 2, 3(a) or 3(b)), 8 and G.1, the bounds $\underline{\Delta_{Y^*}^{OO}}$ and $\overline{\Delta_{Y^*}^{OO}}$, given by equations (13) and (14) are infinitesimally point-wise sharp, i.e., for any $\epsilon \in \mathbb{R}_{++}$, $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x}, \overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x}, \overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x}, \overline{u})\right)$, there exist random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ such that*

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x}, \overline{u}) := \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x}, \overline{u}), \tag{G.1}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{G.2}$$

$$F_{\tilde{V}|X,\tilde{U}} \text{ is a continuous function of } \tilde{U}, \tag{G.3}$$

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}} \text{ is a continuous function of } \tilde{U}, \tag{G.4}$$

and

$$\left|F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y, d, s, z, \overline{x}) - F_{Y,D,S,Z,X}(y, d, s, z, \overline{x})\right| \leq \epsilon \tag{G.5}$$

for any $(y, d, s, z) \in \mathbb{R}^4$, where $\tilde{D} := \mathbf{1}\left\{P(X,Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0,X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1,X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ and $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

Proposition G.3 *Suppose that the functions m_0^Y, m_1^Y, m_0^S and Δ_S are point-identified at every pair $(x, u) \in \mathcal{X} \times [0,1]$. Impose assumptions 1-6, 8 and G.1. If $\mathcal{Y}^* = \mathbb{R}$, then, for any*

$\epsilon \in \mathbb{R}_{++}$, $\overline{u} \in [0,1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x},\overline{u}) \in \mathbb{R}$, there exist random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ such that

$$\Delta_{\tilde{Y}^*}^{OO}(\overline{x},\overline{u}) \coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \middle| X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] = \delta(\overline{x},\overline{u}), \tag{G.6}$$

$$\mathbb{P}\left[\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{V}\right) \in \mathcal{Y}^* \times \mathcal{Y}^* \times [0,1] \middle| X = \overline{x}, \tilde{U} = u\right] = 1 \text{ for any } u \in [0,1], \tag{G.7}$$

$$F_{\tilde{V}|X,\tilde{U}} \text{ is a continuous function of } \tilde{U}, \tag{G.8}$$

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}} \text{ is a continuous function of } \tilde{U}, \tag{G.9}$$

and

$$\left|F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}(y,d,s,z,\overline{x}) - F_{Y,D,S,Z,X}(y,d,s,z,\overline{x})\right| \leq \epsilon \tag{G.10}$$

for any $(y,d,s,z) \in \mathbb{R}^4$, where $\tilde{D} \coloneqq \mathbf{1}\left\{P(X,Z) \geq \tilde{U}\right\}$, $\tilde{S}_0 = \mathbf{1}\left\{Q(0,X) \geq \tilde{V}\right\}$, $\tilde{S}_1 = \mathbf{1}\left\{Q(1,X) \geq \tilde{V}\right\}$, $\tilde{Y}_0 = \tilde{S}_0 \cdot \tilde{Y}_0^*$, $\tilde{Y}_1 = \tilde{S}_1 \cdot \tilde{Y}_1^*$ and $\tilde{Y} = \tilde{D} \cdot \tilde{Y}_1 + \left(1 - \tilde{D}\right) \cdot \tilde{Y}_0$.

The proofs of propositions G.2 and G.3 are below. They are small modification of the previous proofs.

Proof of Proposition G.2. I only prove proposition G.2 under assumption 7.3 (subcases (a) and (b)).The proofs of proposition G.2 under assumptions 7.1 and 7.2 are trivial modifications of the proof presented below.

Fix any sufficiently small $\epsilon \in \mathbb{R}_{++}$, any $\overline{u} \in [0,1]$, any $\overline{x} \in \mathcal{X}$ and any $\delta(\overline{x},\overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u}), \overline{\Delta_{Y^*}^{OO}}(\overline{x},\overline{u})\right)$. For brevity, define $\alpha(\overline{x},\overline{u}) \coloneqq \delta(\overline{x},\overline{u}) + \dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})}$, $\gamma(\overline{x},\overline{u}) \coloneqq \dfrac{m_1^Y(\overline{x},\overline{u}) - \alpha(\overline{x},\overline{u}) \cdot m_0^S(\overline{x},\overline{u})}{\Delta_S(\overline{x},\overline{u})}$ and $\overline{\epsilon} \coloneqq \dfrac{\epsilon}{2 \cdot F_X(\overline{x})}$.

Note that

$$
\begin{aligned}
\delta\left(\overline{x},\overline{u}\right) &\in \left(\underline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right), \overline{\Delta_{Y^*}^{OO}}\left(\overline{x},\overline{u}\right)\right) \\
\Leftrightarrow \quad \alpha\left(\overline{x},\overline{u}\right) &\in \left(\max\left\{\frac{m_1^Y\left(x,u\right)-\overline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}, \underline{y}^*\right\},\right. \\
&\qquad \left.\min\left\{\frac{m_1^Y\left(x,u\right)-\underline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}, \overline{y}^*\right\}\right) \\
&\subseteq \left(\underline{y}^*, \overline{y}^*\right),
\end{aligned}
\tag{G.11}
$$

and that

$$
\begin{aligned}
\alpha\left(\overline{x},\overline{u}\right) &\in \left(\frac{m_1^Y\left(x,u\right)-\overline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}, \frac{m_1^Y\left(x,u\right)-\underline{y}^*\cdot\Delta_S\left(x,u\right)}{m_0^S\left(x,u\right)}\right) \\
\Leftrightarrow \quad \gamma\left(\overline{x},\overline{u}\right) &\in \left(\underline{y}^*, \overline{y}^*\right).
\end{aligned}
\tag{G.12}
$$

The strategy of this proof consists of defining random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ through their joint cumulative distribution function $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ and, then, checking that conditions (G.1)-(G.5) are satisfied. I fix $(y_0, y_1, u, v, z, x) \in \mathbb{R}^6$ and define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}$ in fourteen steps:

Step 1. For $x \notin \mathcal{X}$, $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0, y_1, u, v, z, x\right) = F_{Y_0^*,Y_1^*,U,V,Z,X}\left(y_0, y_1, u, v, z, x\right)$.

Step 2. From now on, assume that $x \in \mathcal{X}$. Since

$$
F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z,X}\left(y_0, y_1, u, v, z, x\right) = F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z\left|x\right.\right)\cdot F_X\left(x\right),
$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z\left|x\right.\right)$. Moreover, I impose

$$
Z \perp\!\!\!\perp \left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)|X
$$

by writing

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0,y_1,u,v,z\,|x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)\cdot F_{Z|X}\left(z\,|x\right),$$

implying that it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 3. For $u\notin[0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)=F_{Y_0^*,Y_1^*,U,V|X}\left(y_0,y_1,u,v\,|x\right)$.

Step 4. From now on, assume that $u\in[0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V}|X}\left(y_0,y_1,u,v\,|x\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)\cdot F_{\tilde{U}|X}\left(u\,|x\right),$$

it suffices to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)$ and $F_{\tilde{U}|X}\left(u\,|x\right)$.

Step 5. I define $F_{\tilde{U}|X}\left(u\,|x\right)=F_{U|X}\left(u\,|x\right)=u$.

Step 6. For any $u\notin(\overline{u}-\overline{\epsilon},\overline{u}+\overline{\epsilon})$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,u\right)=F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,u\right)$.

Step 7. For any $v\notin[0,1]$, I define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,\overline{u}\right)=F_{Y_0^*,Y_1^*,V|X,U}\left(y_0,y_1,v\,|x,\overline{u}\right)$.

Step 8. From now on, assume that $v\in[0,1]$. Since

$$F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{V}|X,\tilde{U}}\left(y_0,y_1,v\,|x,\overline{u}\right)=F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)\cdot F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right),$$

it is sufficient to define $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0,y_1\,|x,\overline{u},v\right)$ and $F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)$.

Step 9. I define

$$F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right)=\begin{cases} m_0^S\left(x,\overline{u}\right)\cdot\dfrac{v}{Q\left(0,x\right)} & \text{if } v\leq Q\left(0,x\right)\\ \\ m_0^S\left(x,\overline{u}\right)+\Delta_S\left(x,\overline{u}\right)\cdot\dfrac{v-Q\left(0,x\right)}{Q\left(1,x\right)-Q\left(0,x\right)} & \text{if } Q\left(0,x\right)<v\leq Q\left(1,x\right)\\ \\ m_1^S\left(x,\overline{u}\right)+\left(1-m_1^S\left(x,\overline{u}\right)\right)\dfrac{v-Q\left(1,x\right)}{1-Q\left(1,x\right)} & \text{if } Q\left(1,x\right)<v\end{cases}.$$

Step 10. For any $u \in (\overline{u} - \overline{\epsilon}, \overline{u})$, I define

$$F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,u\right) = F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}-\overline{\epsilon}\right) \cdot \left(\frac{\overline{u}-u}{\overline{\epsilon}}\right) + F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right) \cdot \left(\frac{u-\overline{u}+\overline{\epsilon}}{\overline{\epsilon}}\right),$$

which are valid cumulative distribution functions because a convex combination of cumulative distribution functions is a cumulative distribution function.

For any $u \in (\overline{u}, \overline{u} + \overline{\epsilon})$, I define

$$F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,u\right) = F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}\right) \cdot \left(\frac{\overline{u}+\overline{\epsilon}-u}{\overline{\epsilon}}\right) + F_{\tilde{V}|X,\tilde{U}}\left(v\,|x,\overline{u}+\overline{\epsilon}\right) \cdot \left(\frac{u-\overline{u}}{\overline{\epsilon}}\right),$$

which are valid cumulative distribution functions because a convex combination of cumulative distribution functions is a cumulative distribution function.

Note that $F_{\tilde{V}|X,\tilde{U}}$ is a continuous function of $\tilde{U}$, i.e., it satisfies restriction (G.3).

Step 11. I write $F_{\tilde{Y}_0^*,\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_0, y_1\,|x,\overline{u},v\right) = F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right) \cdot F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\,|x,\overline{u},v\right)$, implying that I can separately define $F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right)$ and $F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}\left(y_1\,|x,\overline{u},v\right)$.

Step 12. When $\mathcal{Y}^*$ is a bounded interval (sub-case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}\left(y_0\,|x,\overline{u},v\right) = \begin{cases} \mathbf{1}\left\{y_0 \geq \dfrac{m_0^Y\left(\overline{x},\overline{u}\right)}{m_0^S\left(\overline{x},\overline{u}\right)}\right\} & \text{if } v \leq Q\left(0,x\right) \\ ----------- & -------- \\ \mathbf{1}\left\{y_0 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2}\right\} & \text{if } Q\left(0,x\right) < v \end{cases}.$$

When $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_0^*|X,\tilde{U},\tilde{V}}(y_0|x,\overline{u},v) = \begin{cases} 0 & \text{if } y_0 < \underline{y}^* \text{ and } v \leq Q(0,x) \\ 1 - \dfrac{\dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_0 < \overline{y}^* \text{ and } v \leq Q(0,x) \\ 1 & \text{if } \overline{y}^* \leq y_0 \text{ and } v \leq Q(0,x) \\ ---------- & -------------- \\ \mathbf{1}\{y_0 \geq \overline{y}^*\} & \text{if } Q(0,x) < v \end{cases}.$$

which are valid cumulative distribution functions because $\dfrac{m_0^Y(\overline{x},\overline{u})}{m_0^S(\overline{x},\overline{u})} \in \left[\underline{y}^*, \overline{y}^*\right]$.

Step 13. When $\mathcal{Y}^*$ is a bounded interval (case (a) in assumption 7.3), I define

$$F_{\tilde{Y}_1^*|X,\tilde{U},\tilde{V}}(y_1|x,\overline{u},v) = \begin{cases} \mathbf{1}\{y_1 \geq \alpha(\overline{x},\overline{u})\} & \text{if } v \leq Q(0,x) \\ -------- & ----------- \\ \mathbf{1}\{y_1 \geq \gamma(\overline{x},\overline{u})\} & \text{if } Q(0,x) < v \leq Q(1,x) \\ -------- & ----------- \\ \mathbf{1}\left\{y_1 \geq \dfrac{\underline{y}^* + \overline{y}^*}{2}\right\} & \text{if } Q(1,x) < v \end{cases}.$$

When $\overline{y}^* = \max\{y \in \mathcal{Y}^*\}$ and $\underline{y}^* = \min\{y \in \mathcal{Y}^*\}$ (case (b) in assumption 7.3), I define

$$F_{\tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}\left(y_1 \,|x, \overline{u}, v\right) = \begin{cases} 0 & \text{if } y_1 < \underline{y}^* \text{ and } v \leq Q\left(0, x\right) \\ 1 - \dfrac{\alpha\left(\overline{x}, \overline{u}\right) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } v \leq Q\left(0, x\right) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } v \leq Q\left(0, x\right) \\ ---------- & -------------------- \\ 0 & \text{if } y_1 < \underline{y}^* \text{ and } Q\left(0, x\right) < v \leq Q\left(1, x\right) \\ 1 - \dfrac{\gamma\left(\overline{x}, \overline{u}\right) - \underline{y}^*}{\overline{y}^* - \underline{y}^*} & \text{if } \underline{y}^* \leq y_1 < \overline{y}^* \text{ and } Q\left(0, x\right) < v \leq Q\left(1, x\right) \\ 1 & \text{if } \overline{y}^* \leq y_1 \text{ and } Q\left(0, x\right) < v \leq Q\left(1, x\right) \\ ---------- & -------------------- \\ \mathbf{1}\left\{y_1 \geq \overline{y}^*\right\} & \text{if } Q\left(1, x\right) < v \end{cases} .$$

which are valid cumulative distribution functions because of equations (G.11) and (G.12).

Step 14. For any $u \in (\overline{u} - \overline{\epsilon}, \overline{u})$, I define

$$\begin{aligned} F_{\tilde{Y}_0^*, \tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}\left(y_0, y_1 \,|x, u, v\right) &= F_{\tilde{Y}_0^*, \tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}\left(y_0, y_1 \,|x, \overline{u} - \overline{\epsilon}, v\right) \cdot \left(\frac{\overline{u} - u}{\overline{\epsilon}}\right) \\ &+ F_{\tilde{Y}_0^*, \tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}\left(y_0, y_1 \,|x, \overline{u}, v\right) \cdot \left(\frac{u - \overline{u} + \overline{\epsilon}}{\overline{\epsilon}}\right), \end{aligned}$$

which are valid cumulative distribution functions because a convex combination of cumulative distribution functions is a cumulative distribution function.

For any $u \in (\overline{u}, \overline{u} + \overline{\epsilon})$, I define

$$\begin{aligned} F_{\tilde{Y}_0^*, \tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}\left(y_0, y_1 \,|x, u, v\right) &= F_{\tilde{Y}_0^*, \tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}\left(y_0, y_1 \,|x, \overline{u}, v\right) \cdot \left(\frac{\overline{u} + \overline{\epsilon} - u}{\overline{\epsilon}}\right) \\ &+ F_{\tilde{Y}_0^*, \tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}\left(y_0, y_1 \,|x, \overline{u} + \overline{\epsilon}, v\right) \left(\frac{u - \overline{u}}{\overline{\epsilon}}\right), \end{aligned}$$

which are valid cumulative distribution functions because a convex combination of cumulative distribution functions is a cumulative distribution function.

Note that $F_{\tilde{Y}_0^*, \tilde{Y}_1^* | X, \tilde{U}, \tilde{V}}$ is a continuous function of $\tilde{U}$, i.e., it satisfies restriction (G.4).

Having defined the joint cumulative distribution function $F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z, X}$, note that equations (G.11) and (G.12), $\dfrac{m_0^Y(\overline{x}, \overline{u})}{m_0^S(\overline{x}, \overline{u})} \in \left[\underline{y}^*, \overline{y}^*\right]$ and steps 7-14 ensure that equation (G.2) holds.

Now, I show, in three steps, that equation (G.1) holds.

Step 15. Observe that

$$
\begin{aligned}
&\mathbb{E}\left[\tilde{Y}_1^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] \\
&= \mathbb{E}\left[\tilde{Y}_1^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, Q(0, \overline{x}) \geq \tilde{V}\right] \\
&= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q(0, \overline{x}) \geq \tilde{V}\right\} \cdot \tilde{Y}_1^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]}{\mathbb{P}\left[Q(0, \overline{x}) \geq \tilde{V} \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]} \\
&= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q(0, \overline{x}) \geq \tilde{V}\right\} \cdot \mathbb{E}\left[\tilde{Y}_1^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V}\right] \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]}{\mathbb{P}\left[Q(0, \overline{x}) \geq \tilde{V} \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]} \\
&= \frac{\int_0^{Q(0,\overline{x})} \mathbb{E}\left[\tilde{Y}_1^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V} = v\right] \, dF_{\tilde{V}|X,\tilde{U}}(v \,|\, x, \overline{u})}{\mathbb{P}\left[Q(0, \overline{x}) \geq \tilde{V} \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]} \\
&= \frac{\int_0^{Q(0,\overline{x})} \alpha(\overline{x}, \overline{u}) \, dF_{\tilde{V}|X,\tilde{U}}(v \,|\, x, \overline{u})}{\mathbb{P}\left[Q(0, \overline{x}) \geq \tilde{V} \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]} \\
&\qquad \text{by step 13} \\
&= \alpha(\overline{x}, \overline{u}). \qquad \text{(G.13)}
\end{aligned}
$$

Step 16. Notice that

$$
\mathbb{E}\left[\tilde{Y}_0^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right]
$$

$$
\begin{aligned}
&= \mathbb{E}\left[\tilde{Y}_0^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, Q(0, \overline{x}) \geq \tilde{V}\right] \\
&= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q(0, \overline{x}) \geq \tilde{V}\right\} \cdot \tilde{Y}_0^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]}{\mathbb{P}\left[Q(0, \overline{x}) \geq \tilde{V} \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]} \\
&= \frac{\mathbb{E}\left[\mathbf{1}\left\{Q(0, \overline{x}) \geq \tilde{V}\right\} \cdot \mathbb{E}\left[\tilde{Y}_0^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V}\right] \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]}{\mathbb{P}\left[Q(0, \overline{x}) \geq \tilde{V} \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]} \\
&= \frac{\int_0^{Q(0,\overline{x})} \mathbb{E}\left[\tilde{Y}_0^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, \tilde{V} = v\right] \mathrm{d}F_{\tilde{V}|X,\tilde{U}}(v \,|x, \overline{u})}{\mathbb{P}\left[Q(0, \overline{x}) \geq \tilde{V} \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]} \\
&= \frac{\int_0^{Q(0,\overline{x})} \frac{m_0^Y(\overline{x}, \overline{u})}{m_0^S(\overline{x}, \overline{u})} \mathrm{d}F_{\tilde{V}|X,\tilde{U}}(v \,|x, \overline{u})}{\mathbb{P}\left[Q(0, \overline{x}) \geq \tilde{V} \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}\right]} \\
&\qquad \text{by step 12} \\
&= \frac{m_0^Y(\overline{x}, \overline{u})}{m_0^S(\overline{x}, \overline{u})}. \qquad \text{(G.14)}
\end{aligned}
$$

Step 17. Note that

$$
\begin{aligned}
\Delta_{\tilde{Y}^*}^{OO}(\overline{x}, \overline{u}) &\coloneqq \mathbb{E}\left[\tilde{Y}_1^* - \tilde{Y}_0^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] \\
&= \mathbb{E}\left[\tilde{Y}_1^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] \\
&\quad - \mathbb{E}\left[\tilde{Y}_0^* \,\middle|\, X = \overline{x}, \tilde{U} = \overline{u}, \tilde{S}_0 = 1, \tilde{S}_1 = 1\right] \\
&= \alpha(\overline{x}, \overline{u}) - \frac{m_0^Y(\overline{x}, \overline{u})}{m_0^S(\overline{x}, \overline{u})} \\
&\qquad \text{by equations (G.13) and (G.14)} \\
&= \delta(\overline{x}, \overline{u}) \\
&\qquad \text{by the definition of } \alpha(\overline{x}, \overline{u}),
\end{aligned}
$$

ensuring that equation (G.1) holds.

Finally, I show, in four steps, that equation (G.5) holds.

Step 18. Fix $(y, d, s, z) \in \mathbb{R}^4$ arbitrarily and observe that expression (G.5) can be simplified to:

$$\begin{aligned}
&\left|F_{\tilde{Y},\tilde{D},\tilde{S},Z,X}\left(y, d, s, z, \overline{x}\right) - F_{Y,D,S,Z,X}\left(y, d, s, z, \overline{x}\right)\right| \leq \epsilon \\
\Leftrightarrow &\left|F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}\left(y, d, s, z\left|\overline{x}\right.\right) \cdot F_X\left(\overline{x}\right) - F_{Y,D,S,Z|X}\left(y, d, s, z\left|\overline{x}\right.\right) \cdot F_X\left(\overline{x}\right)\right| \leq \epsilon \\
\Leftrightarrow &\left|F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}\left(y, d, s, z\left|\overline{x}\right.\right) - F_{Y,D,S,Z|X}\left(y, d, s, z\left|\overline{x}\right.\right)\right| \leq \frac{\epsilon}{F_X\left(\overline{x}\right)} \\
\Leftrightarrow &\left|F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}\left(y, d, s, z\left|\overline{x}\right.\right) - F_{Y,D,S,Z|X}\left(y, d, s, z\left|\overline{x}\right.\right)\right| \leq 2 \cdot \overline{\epsilon}
\end{aligned} \tag{G.15}$$

by the definition of $\overline{\epsilon}$.

Step 19. Notice that

$$\begin{aligned}
&F_{\tilde{Y},\tilde{D},\tilde{S},Z|X}\left(y, d, s, z\left|\overline{x}\right.\right) - F_{Y,D,S,Z|X}\left(y, d, s, z\left|\overline{x}\right.\right) \\
&\quad= \mathbb{E}\left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\}\middle| X = \overline{x}\right] - \mathbb{E}\left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\}\middle| X = \overline{x}\right] \\
&\quad= \int \mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z\left|\overline{x}\right.\right) \\
&\qquad - \int \mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z\left|\overline{x}\right.\right) \\
&\quad= \int \left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \notin \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z\left|\overline{x}\right.\right) \\
&\qquad + \int \left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \in \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z\left|\overline{x}\right.\right) \\
&\qquad - \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \notin \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z\left|\overline{x}\right.\right) \\
&\qquad - \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \in \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z\left|\overline{x}\right.\right)
\end{aligned}$$

by linearity of the Lebesgue Integral

$$\begin{aligned}
&\quad= \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \notin \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z\left|\overline{x}\right.\right) \\
&\qquad + \int \left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \in \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z\left|\overline{x}\right.\right) \\
&\qquad - \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \notin \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z\left|\overline{x}\right.\right) \\
&\qquad - \int \left[\mathbf{1}\left\{(Y, D, S, Z) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \in \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{Y_0^*,Y_1^*,U,V,Z|X}\left(y_0, y_1, u, v, z\left|\overline{x}\right.\right)
\end{aligned}$$

by steps 2-6

$$\quad= \int \left[\mathbf{1}\left\{\left(\tilde{Y}, \tilde{D}, \tilde{S}, Z\right) \leq (y, d, s, z)\right\} \cdot \mathbf{1}\left\{u \in \left(\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}\right)\right\}\right] \mathrm{d}F_{\tilde{Y}_0^*,\tilde{Y}_1^*,\tilde{U},\tilde{V},Z|X}\left(y_0, y_1, u, v, z\left|\overline{x}\right.\right)$$

$$
\begin{aligned}
&\quad - \int \left[\mathbf{1}\left\{ (Y, D, S, Z) \leq (y, d, s, z) \right\} \cdot \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \right] \mathrm{d}F_{Y_0^*, Y_1^*, U, V, Z|X} \left(y_0, y_1, u, v, z \,|\overline{x} \right) \\
&\leq \int \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \mathrm{d}F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z|X} \left(y_0, y_1, u, v, z \,|\overline{x} \right) \\
&= \int \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \mathrm{d}F_{\tilde{U}|X} \left(u \,|\overline{x} \right) \\
&= 2 \cdot \overline{\epsilon}
\end{aligned}
$$

by step 5.

Step 20. Following the same procedure of step 19, I have that:

$$
\begin{aligned}
&F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z|X} \left(y, d, s, z \,|\overline{x} \right) - F_{Y, D, S, Z|X} \left(y, d, s, z \,|\overline{x} \right) \\
&= \int \left[\mathbf{1}\left\{ \left(\tilde{Y}, \tilde{D}, \tilde{S}, Z \right) \leq (y, d, s, z) \right\} \cdot \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \right] \mathrm{d}F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z|X} \left(y_0, y_1, u, v, z \,|\overline{x} \right) \\
&\quad - \int \left[\mathbf{1}\left\{ (Y, D, S, Z) \leq (y, d, s, z) \right\} \cdot \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \right] \mathrm{d}F_{Y_0^*, Y_1^*, U, V, Z|X} \left(y_0, y_1, u, v, z \,|\overline{x} \right) \\
&\geq - \int \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \mathrm{d}F_{Y_0^*, Y_1^*, U, V, Z|X} \left(y_0, y_1, u, v, z \,|\overline{x} \right) \\
&= - \int \mathbf{1}\left\{ u \in (\overline{u} - \overline{\epsilon}, \overline{u} + \overline{\epsilon}) \right\} \mathrm{d}F_{U|X} \left(u \,|\overline{x} \right) \\
&= -2 \cdot \overline{\epsilon}
\end{aligned}
$$

Step 21. Combining steps 19 and 20, I find that

$$
\left| F_{\tilde{Y}, \tilde{D}, \tilde{S}, Z|X} \left(y, d, s, z \,|\overline{x} \right) - F_{Y, D, S, Z|X} \left(y, d, s, z \,|\overline{x} \right) \right| \leq 2 \cdot \overline{\epsilon},
$$

implying equation (G.5) according to equation (G.15).

I can, then, conclude that proposition G.2 is true. ■

Proof of Proposition G.3. This proof is essentially the same proof of proposition G.2 under assumption 7.3.(a). Fix $\epsilon \in \mathbb{R}_{++}$, $\overline{u} \in [0, 1]$, $\overline{x} \in \mathcal{X}$ and $\delta(\overline{x}, \overline{u}) \in \left(\underline{\Delta_{Y^*}^{OO}} (\overline{x}, \overline{u}), \overline{\Delta_{Y^*}^{OO}} (\overline{x}, \overline{u}) \right)$ arbitrarily. For brevity, define $\alpha(\overline{x}, \overline{u}) := \delta(\overline{x}, \overline{u}) + \dfrac{m_0^Y(\overline{x}, \overline{u})}{m_0^S(\overline{x}, \overline{u})}$, $\gamma(\overline{x}, \overline{u}) := \dfrac{m_1^Y(\overline{x}, \overline{u}) - \alpha(\overline{x}, \overline{u}) \cdot m_0^S(\overline{x}, \overline{u})}{\Delta_S(\overline{x}, \overline{u})}$ and $\overline{\epsilon} := \dfrac{\epsilon}{2 \cdot F_X(\overline{x})}$. Note that $\alpha(\overline{x}, \overline{u}) \in \mathbb{R} = \mathcal{Y}^*$ and $\gamma(\overline{x}, \overline{u}) \in \mathbb{R} = \mathcal{Y}^*$.

I define the random variables $\left(\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}\right)$ using the joint cumulative distribution function $F_{\tilde{Y}_0^*, \tilde{Y}_1^*, \tilde{U}, \tilde{V}, Z, X}$ described by steps 1-14 in the proof of proposition G.2 for the case of convex support $\mathcal{Y}^*$. Note that equation (G.7) is trivially true when $\mathcal{Y}^* = \mathbb{R}$. Moreover, equations (G.6) and (G.10) are valid by the argument described in steps 15-21 in the previous proof.

I can, then, conclude that proposition G.3 is true. ■